Carlos Ivorra Castillo

INTRODUCCIÓN A LA TEORÍA ALGEBRAICA DE NÚMEROS

It is ordinary rational arithmetic which attracts the ordinary man.

G.H. Hardy

Índice General

Preámbulo

Una de las características de la teoría de números es que muchos de los problemas que aborda requieren para su resolución de conocimientos más profundos que los necesarios para entender el problema. Esto hace que haya muchos aficionados a los números que traten de enfrentarse a este tipo de problemas sin apenas más base que el álgebra y la aritmética elemental, lo cual se traduce en que, o bien no consiguen resultados, o bien obtienen demostraciones técnicas y farragosas considerando incluso como mérito el haber evitado técnicas más sofisticadas, aunque éstas sean mucho más elegantes e iluminadoras.

Cuando alguno de estos aficionados a los números, movido por la curiosidad, trata de enfrentarse a un libro que podría proporcionarle técnicas más potentes para abordar los problemas que le interesan, es fácil que tropiece con la barrera que suponen los requisitos de álgebra abstracta o análisis matemático que tales libros presuponen.

Este libro pretende hacer accesibles a un matemático aficionado sin base alguna de "matemática abstracta" muchos resultados de la teoría algebraica de números que, con las exposiciones habituales, requieren ciertos conocimientos sobre anillos, grupos, módulos, etc.

Para ello, no supondremos en el lector ningún conocimiento previo más allá de cierta familiaridad y soltura con el álgebra y la aritmética más elementales.

Para precisar el perfil del tipo de lector al que nos estamos refiriendo, hablamos de un lector que pueda abordar (en el sentido de resolver o, al menos, de entender la solución y verla natural) un problema como éste:

> *Determinar los números naturales que pueden expresarse como diferencia de dos cuadrados, es decir, $n = x^2 - y^2$, donde x e y son números naturales.*

Aunque este problema es muy simple, el lector debería adquirir el hábito de experimentar antes de teorizar. Los grandes maestros de la teoría de números, como Fermat, Euler, Gauss, etc., sólo obtuvieron demostraciones generales de muchos hechos después de haberlos conjeturado a partir de cálculos particulares. En muchas ocasiones pudieron conjeturar mucho más de lo que luego fueron capaces de demostrar. Y la perspicacia necesaria para formular una conjetura correcta a partir de unos cálculos particulares no es inferior en muchos casos al ingenio que requiere encontrar una prueba general.

Lo ideal sería que el lector dispusiera de un ordenador al que pudiera encargarle cálculos sencillos rutinarios, como por ejemplo, calcular $x^2 - y^2$ para todos los valores de x, y, digamos entre 0 y 100, ordenar los resultados obtenidos e imprimir los primeros, por ejemplo, los valores entre 0 y 100. Si lo hace, el resultado que obtendrá será éste:

0	1	3	4	5	7	8	9	11	12	13	15	16	17	19
20	21	23	24	25	27	28	29	31	32	33	35	36	37	39
40	41	43	44	45	47	48	49	51	52	53	55	56	57	59
60	61	63	64	65	67	68	69	71	72	73	75	76	77	79
80	81	83	84	85	87	88	89	91	92	93	95	96	97	99

A partir de aquí es muy fácil llegar esta conjetura:

> *Un número entero es de la forma $x^2 - y^2$ si y sólo si es impar o múltiplo de 4.*

A la hora de probarlo el lector debería considerar como un paso natural recurrir a esta fórmula:

$$n = x^2 - y^2 = (x + y)(x - y).$$

Dado un número n, buscamos dos números x e y que cumplan esto. Como no sabemos nada sobre los posibles divisores de n, podemos tratar de elegir x, y de modo que, por ejemplo,

$$x + y = n, \qquad x - y = 1.$$

Para que se cumpla esto, tiene que ser $x = y + 1$ y, al sustituir esta ecuación en la primera, queda $2y + 1 = n$. Vemos que, para que este camino dé resultado, es necesario que n sea impar. Dicho al revés: si partimos de un número impar $n = 2k + 1$, vemos que basta tomar $y = k$ y entonces $x = k + 1$. En resumen:

$$2k + 1 = (k + 1)^2 - k^2.$$

¿Qué hacemos si $n = 2m$ es par? Entonces x, y tienen que ser ambos pares o ambos impares, luego no podemos pretender que $x - y = 1$. La posibilidad más simple en este caso es ver si podemos hacer que

$$x + y = m, \quad x - y = 2.$$

Para esto tiene que ser $x = y + 2$ y la primera ecuación nos da que $2y + 2 = m$, luego necesitamos que $m = 2k$ sea par (es decir, que n sea múltiplo de 4, tal y como habíamos conjeturado). Si lo suponemos así, necesitamos que $2y+2 = 2k$, luego $y = k-1$, y entonces $x = y+2 = k-1+2 = k+1$. En resumen, llegamos a la expresión

$$4k = (k + 1)^2 - (k - 1)^2.$$

Con esto hemos probado que todos los números impares y todos los múltiplos de 4 se pueden expresar como diferencia de dos cuadrados, pero falta demostrar

que los números pares que no son múltiplos de 4, es decir, los de la forma $n = 2m$ con $m = 2k + 1$ impar o, equivalentemente, $n = 4k + 2$, no pueden expresarse como diferencia de dos cuadrados.

Para ello, no es imprescindible, pero sí útil, recurrir al concepto de congruencia. La tabla siguiente muestra los cuatro restos posibles módulo 4 de un número x y los restos correspondientes a x^2:

x	0	1	2	3
x^2	0	1	0	1

Concluimos que todo cuadrado cumple $x^2 \equiv 0, 1 \pmod 4$, luego una diferencia de dos cuadrados cumple

$$n = x^2 - y^2 \equiv 0, 1, -1 \pmod 4,$$

pero es imposible $n \equiv 2 \pmod 4$, es decir, es imposible que $n = 2k + 4$ si n es diferencia de dos cuadrados. ∎

Tal vez el lector no esté familiarizado con la noción de congruencia, pero no importa, porque la estudiaremos con detalle en el capítulo III. Las congruencias suelen estar en la frontera entre lo que los aficionados[1] a la teoría de números con reticencias hacia el álgebra abstracta llegan a "interiorizar" espontáneamente o no. Para no dejar a ningún lector atrás por falta de base suficiente, dedicaremos los primeros capítulos a revistar todos los conocimientos necesarios prácticamente desde cero.

No hay que deducir de aquí que en este libro vayamos a considerar únicamente razonamientos elementales como los que hemos empleado en el ejemplo precedente. Todo lo contrario: afirmamos que no supondremos al lector más conocimientos que los necesarios para seguir esa clase de razonamientos, pero el lector que siga este libro hasta el final llegará a familiarizarse con conceptos algebraicos bastante abstractos. Esencialmente, lo que vamos a ver es cómo muchos problemas sobre números enteros pueden concebirse y abordarse más adecuadamente en términos los llamados "anillos de enteros algebraicos".

Por ejemplo, consideremos lo que aparentemente es una pequeña variante del problema que hemos analizado antes:

> *Determinar los números naturales que pueden expresarse como suma de dos cuadrados, es decir, $n = x^2 + y^2$, donde x e y son números naturales.*

[1] La palabra "aficionado" tiene a veces un matiz despectivo que bajo ningún concepto debe presuponerse en el uso que aquí le estamos dando. Obviamente, entre los aficionados a la teoría de números —como en cualquier otro grupo de personas— los hay de todas las clases, incluyendo quienes no dudan en hacer gala de la más recalcitrante ignorancia para encadenar una barbaridad detrás de otra, pero también mentes inquietas, con una curiosidad y un deseo de aprender dignos de elogio, que a menudo dan muestras de un ingenio, una inteligencia y una perspicacia poco frecuentes incluso entre estudiantes "profesionales". Es obviamente a éstos a quienes está dirigido este libro y de ellos estamos hablando.

De nuevo el lector está invitado a investigar empíricamente el problema, calculando una lista, digamos, con todos los números menores que 100 con esta propiedad y tratando de conjeturar un criterio sencillo que determine cuál es están y cuáles no están en ella. En este caso la conjetura no es tan fácil de obtener. Como pista diremos únicamente que depende de la descomposición en factores primos de n.

Aunque es posible formular y demostrar el criterio mediante técnicas elementales, sucede que es mucho más claro desde un punto de vista conceptual usar la factorización

$$n = (x + yi)(x - yi),$$

donde $i = \sqrt{-1}$ es la "unidad imaginaria" y reformular el problema en términos de los llamados "enteros de Gauss", que son los números de la forma $a+bi$, donde a, b son enteros "ordinarios". En el capítulo IV demostraremos que estos enteros "generalizados" tienen propiedades aritméticas muy similares a las de los enteros "ordinarios", incluyendo la posibilidad de dividirlos euclídeamente (con cociente y resto) y esto permite aplicar los resultados del capítulo II para concluir que todo entero de Gauss admite una descomposición única en factores primos.

Mientras una respuesta "elemental" al problema de las sumas de cuadrados es necesariamente técnica y laboriosa, cuando el lector haya asimilado debidamente la aritmética de los enteros de Gauss, la solución del problema le resultará poco menos que evidente.

Los enteros de Gauss forman un caso particular de anillos algebraicos. En este libro usaremos éstos y otros similares, como los enteros de Eisenstein, y otros más complejos, como los enteros ciclotómicos, para obtener con relativa facilidad resultados de enunciado elemental, que pueden entenderse sin necesidad de conocer tales conceptos. Precisamente porque los enunciados son elementales, los hemos reunido en una lista justo a continuación de este preámbulo para que el lector pueda ojearla y juzgar si le parecen interesantes, e incluso trate de probar algunos de ellos para juzgar su nivel de dificultad.

Al igual que introduciremos los enteros de Gauss, paulatinamente iremos introduciendo más conceptos algebraicos, a menudo en casos particulares, sin ninguna vocación de generalidad, y sólo en la medida en que vayan siendo necesarios para abordar problemas concretos planteados previamente. Igualmente reduciremos al mínimo la teoría necesaria. Por ejemplo, el álgebra lineal que usaremos se reducirá a las propiedades básicas de las matrices y determinantes de orden 2×2, la teoría de grupos se reducirá a algunas propiedades elementales sobre el orden de un elemento de un grupo finito, etc.

No quisiéramos dar a entender con esto ninguna clase de desprecio por los conceptos matemáticos abstractos. Al contrario, esperamos que este libro pueda ayudar a algunos aficionados a la teoría de números a apreciar las posibilidades que ofrece lo que se conoce como el "álgebra abstracta", pues, si bien es cierto que vamos a ver muy poca, esperamos que sea la suficiente como para que el lector comprenda que, incluso aprendiendo todo lo que puede aprender de este libro, se encontrará nadando en la orilla de un océano en el que uno no se puede adentrar sin una embarcación adecuada, y esa embarcación es precisamente el álgebra abstracta.

La diferencia entre exponer los conceptos algebraicos en toda su generalidad o trabajar con casos particulares como vamos a hacer aquí puede equipararse a la diferencia entre estudiar geometría en espacios de dimensión arbitraria n o estudiar la geometría plana. Como introducción a la geometría puede ser útil tratar exclusivamente el caso bidimensional, si bien éste se quedará corto a la hora de abordar muchos problemas que no requieren realmente ideas nuevas, sino la mera generalización a dimensiones arbitrarias. Tal vez el lector que se familiarice con el uso de técnicas algebraicas en el caso "bidimensional" que aquí vamos a tratar encuentre en ellas la motivación necesaria para apreciar su estudio en general.

De hecho, esperamos incluso que este libro pueda resultar motivador e iluminador para un estudiante de matemáticas que esté empezando a tratar con anillos, ideales, cocientes, grupos, determinantes, etc. sin tener una idea clara de la finalidad de tales conceptos, pues aquí encontrará una "versión en miniatura" —pero nada trivial— de lo que puede lograrse con la aplicación sistemática de tales conceptos.

Este libro está pensado para que pueda ser compaginado con mi libro de *Introducción a la teoría analítica de números* (en lo sucesivo [ITAn]). Por regla general, serán más los conceptos y resultados introducidos aquí que se usarán también en [ITAn] que viceversa. Entre los resultados que "importaremos" de [ITAn] se encuentran el teorema fundamental del álgebra, y en particular la existencia de raíces de la unidad en el cuerpo $\mathbb{C}$ de los números complejos y el teorema de Dirichlet sobre primos en progresiones aritméticas, del que no se conoce una demostración sin elementos analíticos. En los últimos capítulos de ambos libros combinaremos buena parte de las técnicas algebraicas y analíticas estudiadas previamente, hasta entonces con bastante independencia. La tabla siguiente muestra el orden en que pueden alternarse los capítulos:

[ITAl]		[ITAn]	
I	Álgebra básica		
II	Aritmética básica	**I**	Sucesiones
III	Congruencias	**II**	Series infinitas
IV	Enteros de Gauss	**III**	Continuidad
V	Símbolo de Legendre	**IV**	La función exponencial
VI	Enteros de Eisenstein	**V**	Funciones trigonométricas
VII	Reciprocidad cuadrática	**VI**	Funciones elementales
VIII	Enteros algebraicos	**VII**	La distribución de los primos
IX	Cuerpos cuadráticos	**VIII**	Series de Dirichlet
X	Fracciones continuas	**IX**	Fracciones continuas
XI	Formas cuadráticas	**X**	Números trascendentes
XII	Módulos		
XIII	Aritmética ideal	**XI**	La función dseta de Dedekind
XIV	Géneros	**A**	La medida de Jordan
XV	Reciprocidad cúbica		
XVI	Reciprocidad bicuadrática		
XVII	Enteros ciclotómicos		

Centrándonos en los contenidos de este libro, en el capítulo I repasaremos el álgebra elemental con la que el lector debería estar familiarizado, es decir, las propiedades básicas de los números naturales, enteros, racionales y reales, así como de los polinomios. También se introducirá algo de vocabulario algebraico que es posible que algunos lectores no conozcan (anillo, dominio íntegro, cuerpo, etc.), pero haremos poco más que introducir las definiciones a modo de nombres descriptivos de "paquetes de propiedades" que nos permitan expresar más cómodamente las diferencias entre los distintos conjuntos de números (y de objetos parecidos a números) que vamos a manejar.

En el capítulo II repasaremos la aritmética básica (descomposición en factores primos, máximo común divisor, mínimo común múltiplo, etc.) mostrando que muchos de los resultados que el lector conocerá sin duda en el caso concreto de los números naturales y enteros son válidos igualmente en otros anillos que cumplan unos requisitos adecuados (esencialmente, en los que tenga sentido realizar divisiones euclídeas, con cociente y resto). Este hecho va a ser fundamental en buena parte del libro.

El capítulo III está dedicado a las congruencias, que, según señalábamos, suele estar en el límite de los conocimientos de los aficionados a la teoría de números sin formación en álgebra abstracta (los hay que las desconocen, los hay que las conocen, pero las manejan "con pinzas" y, por supuesto, también los hay que las dominan). Aquí aprovecharemos para introducir unos pocos conceptos elementales de la teoría de grupos finitos, esencialmente el concepto de orden de un elemento y el hecho de que el orden divide al número de elementos del grupo. Veremos que este hecho elemental tiene aplicaciones en los contextos más diversos.

En los capítulos siguientes presentamos paulatinamente ideas que, en general, serán novedosas para los lectores que no han pasado la barrera del "algebra abstracta".

En la exposición hemos evitado en lo posible tecnicismos innecesarios habituales en los libros de matemáticas, incluso en muchos considerados "elementales", y que pueden incrementar la altura de la "barrera" que mantiene alejados a los lectores potenciales sin unos conocimientos previos suficientes. Por ejemplo, lo habitual en los libros es definir los números enteros como clases de equivalencia de pares de números naturales. Esto es un tecnicismo que simplifica considerablemente las comprobaciones que requiere la construcción de los números enteros, pero que a cambio puede hacer que muchos lectores que tengan bien claro qué son los números enteros desconfíen de un libro que convierte en algo "oscuro" una idea tan elemental. Aquí hemos optado por definir los números enteros como números naturales precedidos de un signo $+$ o $-$.

Por otro lado, es importante señalar que el libro no tiene lagunas, es decir, que no utilizamos ningún resultado que no hayamos demostrado previamente (salvo unos pocos demostrados en [ITAn]). Esto no impide obviamente que citemos a título informativo algunos resultados que no podemos demostrar aquí (como el Último Teorema de Fermat), pero los resultados enunciados sin demostración no son usados en ningún momento.

No quisiéramos dar la impresión con todo lo dicho de que consideramos éste como el gran libro definitivo que abrirá las puertas de la teoría algebraica de números a los aficionados que habían fracasado con otros libros. Posiblemente algunos lectores preferirán otro tipo de libro como introducción y también es probable que algunos lo encuentren demasiado difícil de seguir. Todo esto dependerá de muchos factores subjetivos. Lo único que hemos querido enfatizar es que, al margen de las preferencias que cada cual pueda tener sobre cuál es la mejor forma de motivar, de exponer o de explicar los temas tratados, este libro ha sido diseñado con la intención de que pueda seguirse sin necesidad de ningún conocimiento teórico previo y de modo que cada resultado teórico se presente en un momento en que pueda entenderse su utilidad a la hora de abordar problemas concretos, pero siempre sin renunciar a justificar razonadamente todos los resultados empleados.

Enunciados elementales probados en este libro

Enumeramos a continuación 126 resultados probados en este libro que admiten un enunciado elemental, es decir, que no requiere emplear conceptos algebraicos más allá del concepto de congruencia. Se indica junto a cada uno el número de página donde está la solución (o donde está planteado como ejercicio). A ellos hay que añadir el problema del ganado de Arquímedes, formulado en la página xxxv, que es demasiado largo para incluirlo aquí.

1. Sea N el producto de todos los 19 números

   $$214\,748\,364\,800\,000, 11, 19, 43, 61, 83, 169, 223, 331, 379, 601, 757, 961,$$

   $$1\,201, 7\,019, 823\,543,\ 616\,318\,177, 6\,561, 100\,895\,598\,169.$$

   y sea P la suma de sus divisores propios. Probar que N divide a P y encontrar el cociente. Determinar si el último número de la lista es primo o compuesto. xxviii

2. Todo número racional puede expresarse como suma de fracciones unitarias (con numerador 1) y denominadores distintos dos a dos. 44

3. Un número natural n es triangular (es de la forma $n = 1 + 2 + \cdots + k$) si y sólo si $8n + 1$ es un cuadrado perfecto. 141

4. Encontrar el menor número natural con 60 divisores (positivos). 87

5. Encontrar el menor número natural con 100 divisores (positivos). 87

6. Encontrar el menor número natural con 5 040 divisores (positivos). 87

7. Encontrar el menor número natural con 3^{10} divisores (positivos). 88

8. El polinomio $f(x) = x^2 + x + 41$ toma valores primos sobre todos los números naturales entre 0 y 39. (Dar una prueba conceptual.) 484

9. Si p es un primo impar, entonces

   $$G(p) = \sum_{k=1}^{p} e^{2k^2\pi i/p} = \begin{cases} \sqrt{p} & \text{si } p \equiv \ \ 1 \pmod 4, \\ \sqrt{p}\,i & \text{si } p \equiv -1 \pmod 4. \end{cases}$$

   258

Ecuaciones diofánticas linales

10. Cinco marineros llegan a una isla desierta y durante el día recogen todos los cocos que pueden encontrar, para asegurarse de que no les faltará agua. Luego acuerdan que al día siguiente los repartirán a partes iguales. Sin embargo, durante la noche, uno de los marineros, temiendo que el reparto no vaya a ser equitativo, decide llevarse su parte en secreto: divide los cocos en cinco montones iguales y se encuentra con que le sobra uno, arroja a los monos el que le sobra y se lleva la quinta parte. Poco después, otro de los marineros tiene la misma idea y de nuevo reparte los cocos en cinco partes iguales, se encuentra con que le sobra uno, lo arroja a los monos y se lleva la quinta parte. Lo mismo hacen, sucesivamente, los tres marineros restantes. Todos ellos se encuentran con que sobra un coco al hacer el reparto y lo arrojan a los monos. A la mañana siguiente, sin que ninguno confiese que ya se ha llevado una parte de las provisiones, los cocos restantes se dividen en cinco partes iguales (esta vez no sobra ninguno) y cada marinero se queda con una de ellas. ¿Cuántos cocos habían recogido los marineros? 74

11. Un hombre cobra un cheque, pero el cajero que se lo paga se equivoca y le da tantos euros como céntimos tendría que haberle dado y tantos céntimos como euros (por ejemplo, si el importe del cheque hubiera sido de de 60.45 euros, el cajero le habría pagado 45.60 euros). Después de haberse gastado 5 céntimos, el hombre advierte el error, pues comprueba que tiene justo el doble de dinero que debería haber cobrado. ¿Cuál era el importe del cheque? 75

Ternas pitagóricas

12. **Clasificación de las ternas pitagóricas** Las soluciones enteras de la ecuación $x^2 + y^2 = z^2$ son (salvo el orden de x e y) los múltiplos de las de la forma
$$(x, y, z) = (p^2 - q^2, 2pq, p^2 + q^2),$$
donde $q < p$ son naturales primos entre sí de paridad opuesta. 76, 186

13. Dados números enteros que cumplan $x^2 + y^2 = z^2$, uno de los tres es múltiplo de 3, uno múltiplo de 4 y uno múltiplo de 5. 115

Casos particulares y variantes del Último Teorema de Fermat

14. La ecuación, $x^4 + y^4 = z^2$ no tiene soluciones enteras no triviales (con todas las variables no nulas). 79

15. La ecuación $x^3 + y^3 = z^3$ no tiene soluciones no triviales (es decir, con todas las variables no nulas). 226

16. La ecuación $x^5 + y^5 = z^5$ no tiene soluciones no triviales (es decir, con todas las variables no nulas). 606

17. La ecuación $x^3 + y^3 = 2z^3$ no tiene soluciones enteras con $x \neq \pm y$. 119

71. **Teorema de Wilson** Si p es primo, entonces $(p-1)! \equiv -1 \pmod{p}$. 135

72. Si p es primo y $p \nmid a$, entonces a tiene raíz n-sima módulo p si y sólo si $a^{(p-1)/d} \equiv 1 \pmod{p}$, donde $d = (p-1, n)$, y entonces tiene exactamente d raíces n-simas módulo p. 164

73. La ecuación $x^2 + 1\,848y^2 = 18\,518\,809$ tiene como única solución (en los números naturales) $197^2 + 1\,848 \cdot 100^2 = 18\,518\,809$. 156

74. La ecuación $x^2 + 357y^2 = 142\,969$ tiene como única solución (en los números naturales) $(x, y) = (13, 20)$. 158

75. Deducir de los dos apartados anteriores que los números 18 518 809 y 142 969 son primos. 518

76. Encontrar los números naturales N tales que el número que resulta de trasladar su primera cifra por la izquierda a la derecha (p.ej. $3\,458 \to 4\,583$) es 1.5 veces mayor que N. 158

77. Encontrar todos los pares de números naturales (x, y) tales que y tiene dos cifras y el producto xy es el número que resulta de anteponer a x a segunda cifra de y y de posponer la primera. 159

Primos de Mersenne

78. Un número par es perfecto (es la suma de sus divisores propios) si y sólo si es de la forma $2^{n-1}(2^n - 1)$ y $2^n - 1$ es primo. 143

79. Si $2^p - 1$ es primo, entonces p también lo es. 144

80. Si p y $2p + 1$ son primos y $p \equiv -1 \pmod{4}$, entonces $2p + 1 \mid 2^p - 1$, por lo que $M_p = 2^p - 1$ no es primo, salvo si $p = 3$. 207

81. Si p, q son primos impares y $q \mid M_p = 2^p - 1$, entonces $q \equiv \pm 1 \pmod{8}$. 208

82. $M_{31} = 2^{31} - 1 = 2\,147\,483\,647$ es primo. 208

83. Un número de la forma $n^k - 1$ con $n > 2$ no es primo. 146

84. **Test de Lucas-Lehmer** Consideremos la sucesión dada por

$$s_0 = 4, \qquad s_{n+1} = s_n^2 - 2.$$

Si p es un primo impar, el número de Mersenne $M_p = 2^p - 1$ es primo si y sólo si $M_p \mid s_{p-2}$. 313

Primos de Fermat

85. Si $2^n + 1$ es primo, entonces n es potencia de 2. 206

86. El número $F_5 = 2^{2^5} + 1 = 4\,294\,967\,297$ no es primo. 206

87. **Test de Pépin** El número de Fermat $F_n = 2^{2^n} + 1$ es primo si y sólo si $3^{(F_n-1)/2} \equiv -1 \pmod{F_n}$. 228

Representación por formas cuadráticas

88. Caracterizar los números que son suma de dos cuadrados. 168, 184

89. Determinar de cuántas formas se puede expresar un número natural como suma de dos cuadrados. 186

90. Caracterizar los números naturales de la forma $n = x^2 + 2y^2$. 198

91. Caracterizar los números naturales de la forma $n = x^2 + 3y^2$. 225

92. Caracterizar los números naturales de la forma $n = x^2 + 5y^2$. 485

93. Caracterizar los números naturales que pueden expresarse en la forma $x^2 - xy + y^2$. Dar una fórmula explícita que muestre que el producto de dos números de esta forma es también de esta forma. 220

94. Caracterizar los números naturales de la forma $x^2 - 2y^2$. 307

95. Caracterizar los números naturales de la forma $x^2 - 3y^2$. 308

96. Caracterizar los números naturales de la forma $x^2 - 5y^2$. 309

97. Caracterizar los primos de la forma $x^2 - xy - 13y^2$ y los de la forma $x^2 - 53y^2$. 403

98. Caracterizar los primos de la forma $x^2 - 14y^2$. 404

99. Un primo p es de la forma:

$p = x^2 - 30y^2$ si y sólo si $p \equiv -29, 1, 19, 49$ (mód 120),
$p = 2x^2 - 15y^2$ si y sólo si $p \equiv -37, -13, -7, 17$ (mód 120) o $p = 2, 3$,
$p = 15x^2 - 2y^2$ si y sólo si $p \equiv -17, 7, 13, 37$ (mód 120),
$p = 30x^2 - y^2$ si y sólo si $p \equiv -49, -19, -1, 29$ (mód 120) o $p = 5$. 510

100. Caracterizar en términos de congruencias los primos que pueden expresarse en la forma $x^2 + 30y^2$, $2x^2 + 15y^2$, $3x^2 + 10y^2$ y $5x^2 + 6y^2$. 510

101. Un entero $m = 3^i \cdot 5^j \cdot m^*$ (donde $(m^*, 15) = 1$) es de la forma $x^2 - 15y^2$ si y sólo si

1) Los primos que lo dividen con exponente impar son $p = 2, 3, 5$ o bien $p \equiv \pm 1, \pm 7, \pm 11, \pm 17$ (mód 60) y

2) Se da uno de los cuatro casos siguientes:

2a) i, j son pares y $m^* \equiv 1, 4$ (mód 15),
2b) i, j son impares y $m^* \equiv -1, -4$ (mód 15),
2c) i es impar, j es par y $m^* \equiv -2, 7$ (mód 15),
2d) i es par, j es impar y $m^* \equiv 2, -7$ (mód 15). 497

102. Un entero no nulo $m = k^2 r$ (donde r es libre de cuadrados) es de la forma $m = 2x^2 - 15y^2$ los primos que dividen a r cumplen $p = 2, 3, 5$ o

$$p \equiv \pm 1, \pm 7, \pm 13, \pm 17, \pm 19, \pm 29, \pm 37, \pm 49 \text{ (mód 120)}.$$

y además se da uno de los casos siguientes:

(a) $(r, 15) = 1$ y $r \equiv 2, 8$ (mód 15),

(b) $3 \mid r$, $5 \nmid r$ y $r/3 \equiv 1, 4$ (mód 15),

(c) $3 \nmid r$, $5 \mid r$ y $r/5 \equiv 7, 13$ (mód 15),

(d) $15 \mid r$ y $r/15 \equiv 11, 14$ (mód 15). 514

103. Un primo p es de la forma $p = 5x^2 + 2xy + 13y^2$ si y sólo si cumple que $p \equiv 5$ (mód 8). 516

104. Todo primo $p \equiv 1$ (mód 3) se expresa en la forma $p = (L^2 + 27M^2)/4$, donde L y M están unívocamente determinados salvo el signo. 527

105. Todo número natural es suma de cuatro cuadrados. 116

106. Un número natural es suma de tres cuadrados si y sólo si no es de la forma $4^k(8l + 7)$. 250

107. Todo número natural es suma de tres números triangulares. 252

Restos cuadráticos

108. **Ley de reciprocidad cuadrática** Si p y q son primos impares y uno de ellos es congruente con 1 módulo 4, entonces p es un resto cuadrático módulo q si y sólo si q es un resto cuadrático módulo p. Si ambos son congruentes con -1 módulo 4, entonces p es un resto cuadrático módulo q si y sólo si q no es un resto cuadrático módulo p. 238, 240

109. **Primera ley suplementaria** Si p es un primo impar, entonces -1 es un resto cuadrático módulo p si y sólo si $p \equiv 1$ (mód 4). 164

110. **Segunda ley suplementaria** Si p es un primo impar, 2 es un resto cuadrático módulo p si y sólo si $p \equiv \pm 1$ (mód 8). 201

111. -2 es un resto cuadrático módulo un primo impar p si y sólo si cumple $p \equiv 1, 3$ (mód 8). 201

112. 3 es un resto cuadrático módulo un número primo $p > 3$ si y sólo si cumple $p \equiv \pm 1$ (mód 12). 200

113. -3 es un resto cuadrático módulo un número primo $p > 3$ si y sólo si cumple $p \equiv 1$ (mód 3). 165

114. Si p es un primo impar y $(a, p) = 1$, entonces a es un resto cuadrático módulo p^k si y sólo si es un resto cuadrático módulo p. 243

(d) 3 es un resto bicuadrático módulo p si y sólo si $4 \mid a-1$ y $3 \mid b$ o bien $4 \mid a+1$ y $3 \mid a$.

(e) 5 es un resto bicuadrático módulo p si y sólo si $5 \mid b$.

(f) -5 es un resto bicuadrático módulo p si y sólo si $4 \mid a-1$ y $5 \mid b$ o bien $4 \mid a+1$ y $5 \mid a$

(g) -7 es un resto bicuadrático módulo p si y sólo si $7 \mid ab$.

(h) 7 es un resto bicuadrático módulo p si y sólo si $4 \mid a-1$ y $7 \mid ab$ o bien $4 \mid a+1$ y $7 \mid a^2-b^2$.

Introducción

Los resultados básicos de la aritmética de los números naturales eran ya conocidos de forma sistemática por los antiguos griegos. Los libros 7–10 de los *Elementos* de Euclides, escritos en Alejandría sobre 300 a.C., constituyen una magnífica exposición de la aritmética básica. Así, por ejemplo, en el libro 7 introduce los números primos y presenta lo que hoy se conoce como algoritmo de Euclides para encontrar el máximo común divisor de dos números naturales. En el libro 9 demuestra que existen infinitos números primos, y en el libro 10 demuestra la irracionalidad de las raíces cuadradas de todos los números naturales que no son cuadrados perfectos.

Pero los griegos se percataron también de que los números satisfacen muchas propiedades "curiosas" que no tienen ninguna utilidad práctica, pero que llaman la atención. Por ejemplo, en la *Metafísica* de Aristóteles (un poco anterior a Euclides) aparecen las fórmulas

$$1+3=2^2, \quad 1+3+5=3^2, \quad 1+3+5+7=4^2, \ldots$$

En el libro 9 de los *Elementos,* Euclides estudia los llamados *números perfectos,* que son los números iguales a la suma de sus divisores propios, como

$$6=1+2+3, \quad 28=1+2+4+7+14, \quad \ldots$$

Euclides demostró que si 2^n-1 es un número primo, entonces $2^{n-1}(2^n-1)$ es un número perfecto, y sólo pudo encontrar tres ejemplos: 6, 28 y 496. Mucho más tarde, en el siglo II d.C., el neopitagórico Nicómaco de Gerasa descubrió uno más: 8 128. También se atribuye a Nicómaco otra curiosa fórmula aritmética:

$$1^3+2^3+3^3+\cdots+n^3=(1+2+3+\cdots+n)^2.$$

El desafío de Mersenne Algunos problemas aritméticos conceptualmente muy simples, se vuelven complicados cuando se aplican a números demasiado grandes para hacer cálculos explícitos. Así, en 1643, el sacerdote francés Marin Mersenne planteo un problema a un juez, también francés, llamado Pierre de Fermat. Mersenne era más cuidadoso que Fermat conservando su correspondencia, así que lo que se conserva es la carta de ese mismo año en la que Fermat le da su solución:

> *Así que usted me pregunta qué proporción tiene el número que se produce de los números siguientes con sus partes alícuotas:*
>
> $$214\,748\,364\,800\,000, 11, 19, 43, 61, 83, 169, 223, 331, 379, 601, 757, 961,$$
>
> $$1\,201, 7\,019, 823\,543,\ 616\,318\,177, 6\,561, 100\,895\,598\,169.$$
>
> *A continuación me pregunta usted si el último número es primo o no, y un método para descubrir en el plazo de un día si es primo o compuesto.*
>
> *A la primera pregunta, yo le respondo que el número que resulta de todos los números precedentes multiplicados entre ellos es subquíntuplo de sus partes.*
>
> *A la segunda pregunta, yo le respondo que el último de estos números es compuesto, y resulta del producto de estos dos:* $898\,493$ *y* $112\,303$.

Traducido a un lenguaje más moderno, el problema planteado por Mersenne es el siguiente:

> *Sea* N *el producto de todos los* 19 *números indicados y sea* P *la suma de sus divisores propios* (una "parte alícuota" de un número N es un divisor de N distinto del propio N).
>
> *Probar que* N *divide a* P *y encontrar el cociente.*
>
> *Determinar si el número* $100\,895\,598\,169$ *es primo o compuesto.*

La respuesta de Fermat es que $P = 5N$ y que

$$100\,895\,598\,169 = 898\,493 \cdot 112\,303.$$

Hoy en día existen programas de cálculo simbólico que permiten resolver este problema en apenas el mismo tiempo que cuesta introducir los datos. Sin embargo, lo interesante es averiguar qué procedimiento pudo haber seguido Fermat para resolverlo "en el plazo de un día". Nadie puede saberlo a ciencia cierta, pues Fermat nunca explicó cómo había llegado a la solución. No obstante, vamos a mostrar una solución completamente elemental que probablemente será bastante similar en esencia al camino que siguió. En primer lugar conviene definir $\sigma(N)$ como la suma de todos los divisores de N (incluyendo el propio N). La tabla muestra los primeros valores de la función σ:

1	1	2	3	3	4	4	7	5	6	6	12	7	8	8	15	9	13	10	18
11	12	12	28	13	14	14	24	15	24	16	31	17	18	18	39	19	20	20	42
21	32	22	36	23	24	24	60	25	31	26	42	27	40	28	56	29	30	30	72
31	32	32	63	33	48	34	54	35	48	36	91	37	38	38	60	39	56	40	90
41	42	42	96	43	44	44	84	45	78	46	72	47	48	48	124	49	57	50	93
51	72	52	98	53	54	54	120	55	72	56	120	57	80	58	90	59	60	60	168
61	62	62	96	63	104	64	127	65	84	66	144	67	68	68	126	69	96	70	144
71	72	72	195	73	74	74	114	75	124	76	140	77	96	78	168	79	80	80	186
81	121	82	126	83	84	84	224	85	108	86	132	87	120	88	180	89	90	90	234
91	112	92	168	93	128	94	144	95	120	96	252	97	98	98	171	99	156	100	217

¿Podría el lector conjeturar a partir de ella algunas de sus propiedades?

Observemos, por ejemplo, que

$$\sigma(10) = 18 = 3 \cdot 6, \qquad \sigma(100) = 217 = 7 \cdot 31.$$

La función σ tiene una propiedad que es más frecuente en este tipo de funciones "aritméticas" de lo que uno podría esperar en un principio:

> Una función f definida sobre los números naturales es *multiplicativa* si cuando $(m, n) = 1$, entonces $f(mn) = f(m)f(n)$.

Aquí (m, n) representa el máximo común divisor de m y n. Para probar que la función σ es multiplicativa basta tener en cuenta que si $(m, n) = 1$, entonces cada divisor de mn se expresa de forma única como el producto de un divisor de m y un divisor de n. En efecto, si $a \mid mn$, descomponemos el número a en un producto, el primero de cuyos factores contenga a los factores primos de a que dividen a m y el segundo a los que dividen a n. Entonces

$$\sigma(mn) = \sum_{d|mn} d = \sum_{u|m} \sum_{v|n} uv = \sum_{u|m} u \Big(\sum_{v|n} v \Big) = \Big(\sum_{u|m} u \Big) \sigma(n) = \sigma(m)\sigma(n).$$

No es cierto que la igualdad $\sigma(mn) = \sigma(m)\sigma(n)$ se cumpla en general (como muestra la tabla precedente), pero el hecho de que una función sea multiplicativa en este sentido débil reduce su cálculo al caso de potencias de primo, pues todo número natural se descompone en producto de potencias de primos distintos, y éstas son primas entre sí dos a dos.

Y sucede que, si p es primo, es fácil calcular

$$\sigma(p^n) = 1 + p + \cdots + p^n = \frac{p^{n+1} - 1}{p - 1}.$$

Así, por ejemplo,

$$\sigma(100) = \sigma(2^2)\sigma(5^2) = \frac{2^3 - 1}{2 - 1} \cdot \frac{5^3 - 1}{5 - 1} = 7 \cdot 31 = 217.$$

Sin más "herramientas", podemos abordar el desafío de Mersenne. En primer lugar, los 18 primeros números dados se pueden descomponer fácilmente[2] en factores primos:

$2^{36} \cdot 5^5$	11	19	43	61	83	13^2	223	331
379	601	757	31^2	1 201	7 019	7^7	616 318 177	3^8

Así pues, si llamamos N al producto de los 19 números considerados por Mersenne, resulta que

$$N = 2^{36} \cdot 3^8 \cdot 5^5 \cdot 7^7 \cdot 11 \cdot 13^2 \cdot 19 \cdot 31^2 \cdot 43 \cdot 61 \cdot 83 \cdot 223 \cdot 331 \cdot 379 \cdot 601 \cdot 757 \cdot$$
$$1\,201 \cdot 7\,019 \cdot 616\,318\,177 \cdot 100\,895\,598\,169,$$

donde todos los factores son potencia de primo salvo quizá el último.

[2] En un tratado del jesuita suizo Paul Guldin publicado en 1643 se incluye una tabla con los factores primos de todos los números impares menores que 10 000. En cuanto al número 616 318 177, en 1640 Fermat le había comunicado a Mersenne en una carta la descomposición $2^{37} - 1 = 223 \cdot 616\,318\,177$, y es probable que ya entonces hubiera comprobado que ambos factores son primos. (Véase la página 209.)

Ahora podemos calcular $\sigma(N)$ calculando σ sobre cada factor y multiplicando. Para ello tenemos que comprobar que el último factor no es divisible entre ninguno de los 19 primos que dividen a los factores restantes, lo cual puede hacerse en relativamente poco tiempo. Para los 19 primeros factores el resultado es:

p^k	$\sigma(p^k)$	
2^{36}	$2^{37}-1$	$= 223 \cdot 616\,318\,177$
3^8	9 841	$= 13 \cdot 757$
5^5	3 906	$= 2 \cdot 3^2 \cdot 7 \cdot 31$
7^7	960 800	$= 2^5 \cdot 5^2 \cdot 1\,201$
11	12	$= 2^2 \cdot 3$
13^2	183	$= 3 \cdot 61$
19	20	$= 2^2 \cdot 5$
31^2	993	$= 3 \cdot 331$
43	44	$= 2^2 \cdot 11$
61	62	$= 2 \cdot 31$
83	84	$= 2^2 \cdot 3 \cdot 7$
223	224	$= 2^5 \cdot 7$
331	332	$= 2^2 \cdot 83$
379	380	$= 2^2 \cdot 5 \cdot 19$
601	602	$= 2 \cdot 7 \cdot 43$
757	758	$= 2 \cdot 379$
1 201	1 202	$= 2 \cdot 601$
7 019	7 020	$= 2^2 \cdot 3^3 \cdot 5 \cdot 13$
616 318 177	616 318 178	$= 2 \cdot 7^3 \cdot 898\,423$

Si llamamos $s = \sigma(100\,895\,598\,169)$, que, de momento, no sabemos calcular, tenemos que

$$\sigma(N) = 2^{30} \cdot 3^9 \cdot 5^5 \cdot 7^7 \cdot 11 \cdot 13^2 \cdot 19 \cdot 31^2 \cdot 43 \cdot 61 \cdot 83 \cdot$$

$$223 \cdot 331 \cdot 379 \cdot 601 \cdot 757 \cdot 1\,201 \cdot 898\,423 \cdot 616\,318\,177 \cdot s.$$

Comparando N y $\sigma(N)$ vemos que un factor común de ambos números es

$$M = 2^{30} \cdot 3^8 \cdot 5^5 \cdot 7^7 \cdot 11 \cdot 13^2 \cdot 19 \cdot 31^2 \cdot 43 \cdot$$

$$61 \cdot 83 \cdot 223 \cdot 331 \cdot 379 \cdot 601 \cdot 757 \cdot 1\,201 \cdot 616\,318\,177.$$

Explícitamente:

$$\begin{aligned} N &= M \cdot 2^6 \cdot 7\,019 \cdot 100\,895\,598\,169, \\ \sigma(N) &= M \cdot 3 \cdot 898\,423 \cdot s. \end{aligned}$$

Por lo tanto, se cumplirá que N divide a $\sigma(N)$ si y sólo si existe un k tal que

$$3 \cdot 898\,423 \cdot s = k \cdot 2^6 \cdot 7\,019 \cdot 100\,895\,598\,169.$$

Esto obliga a que k sea múltiplo de 3, y en este punto es natural plantearse si 898 423 divide a 100 895 598 169, y probablemente fue así como Fermat obtuvo que

$$100\,895\,598\,169 = 898\,493 \cdot 112\,303.$$

Ahora es necesario comprobar que ambos factores son primos[3] para calcular

$$s = \sigma(898\,493)\sigma(112\,303) = (2^3 \cdot 112\,303)(2^4 \cdot 7\,019),$$

lo que nos reduce la ecuación que determina k a

$$2^7 \cdot 3 \cdot 7\,019 \cdot 112\,303 \cdot 898\,423 = k \cdot 2^6 \cdot 7\,019 \cdot 898\,493 \cdot 112\,303,$$

que se simplifica hasta $k = 6$. Esto significa que $\sigma(N) = 6N$, pero $\sigma(N)$ es la suma de todos los divisores de N incluyendo a N, por lo que la suma de los divisores propios es

$$\sigma(N) - N = 5N,$$

tal y como Fermat indicó.

Si Fermat siguió un camino similar a éste para resolver el desafío de Mersenne, lo cierto es que no necesitó un "método pare descubrir en el plazo de un día" si el número 100 895 598 169 es primo o compuesto, pues el factor 898 493 viene sugerido por el propio problema.

Vamos ahora a mostrar un método elemental que permite factorizar el número dado sin partir de ninguna información adicional en un tiempo razonable. Se trata de un refinamiento de un método descrito por el propio Fermat en una carta a Mersenne de 1664, así que es probable que figurara entre sus técnicas nunca reveladas.

Partimos de un número N, fijamos arbitrariamente un valor $k \geq 1$ y vamos a buscar dos posibles factores de N bajo el supuesto de que uno de ellos sea aproximadamente k veces mayor que el otro. Más concretamente, suponemos que

$$N = (kq + a)(q + b),$$

donde $q = E[\sqrt{N/k}]$ y a y b son pequeños, digamos $|a|, |b| < \sqrt{q}$. Por simplicidad vamos a suponer que ambos son positivos (por definición de q, no pueden ser ambos negativos). Entonces

$$N - kq^2 = (kb + a)q + ab, \qquad 0 \leq ab < q,$$

luego, si llamamos c y r al cociente y el resto de la división euclídea de $N - kq^2$ entre q, tiene que ser[4]

$$c = kb + a, \qquad r = ab.$$

Si fuera $r = 0$ tenemos que $q \mid N$ y ya tenemos una factorización de N. En otro caso b es una raíz de la ecuación $kb^2 - cb + r = 0$, de discriminante $\Delta = c^2 - 4kr$. Observemos que para que la ecuación tenga solución es necesario que $4kr \leq c^2$.

[3] Por ejemplo, para comprobar que 898 493 es primo basta tratar de dividirlo entre los 161 primos menores que su raíz cuadrada (hasta 947) y para el 112 302 hay que considerar los 67 primos menores o iguales que 331.

[4] En el caso $ab < 0$, que hemos descartado, habría que considerar también la división euclídea con cociente por exceso.

Si la ecuación admite una solución entera $b > 0$ tal que $a = r/b$ también es entero, entonces hemos encontrado una descomposición de N en producto de dos factores no triviales.

Al aplicar este método a $N = 100\,895\,598\,169$ obtenemos lo siguiente:

k	q	$N - kq^2$	c	r
1	317 640	428 569	1	110 929
2	224 605	786 119	3	112 304
3	183 389	1 022 206	5	105 261
4	158 820	428 569	2	110 929
5	142 053	324 124	2	40 018
6	129 676	408 313	3	19 285
7	120 056	1 496 217	12	55 545
8	112 302	1 684 537	15	7

Vemos que, para $k \leq 7$, no se cumple la condición necesaria $r \leq c^2/4k$, mientras que para $k = 8$ la ecuación para b resulta ser

$$8b^2 - 15b + 7 = 0,$$

cuyas raíces son $b = 1$ y $b = 7/8$. Con $b = 1$ obtenemos $a = r/1 = 7$, lo que nos da la descomposición

$$N = (8q + 7)(q + 1) = 898\,423 \cdot 112\,303,$$

donde, en efecto, un factor es aproximadamente 8 veces mayor que el otro.

Números de Mersenne En 1644 Mersenne publicó su tratado *Cogitata physico-mathematica,* en cuyo prólogo, en relación con el teorema de Euclides sobre números perfectos, afirmó que, hasta $p = 257$, los únicos números de la forma $M_p = 2^p - 1$ que son primos son los correspondientes a los valores

$$p = 2, \quad 3, \quad 5, \quad 7, \quad 13, \quad 19, \quad 31, \quad 67, \quad 127, \quad 257.$$

No dio ninguna indicación de cómo había llegado a dicha conclusión, pero el hecho es que ni $2^{67} - 1$ ni $2^{257} - 1$ son números primos y, por el contrario, sí que lo son los números correspondientes a los exponentes $p = 61, 89, 107$.

Mersenne se dio cuenta de algo que, al parecer, a Euclides le había pasado inadvertido, y es que para que M_p pueda ser primo es necesario que p lo sea. Los números de la forma $M_p = 2^p - 1$, con p primo se conocen como *números de Mersenne.* Notemos que, a poco que aumentan los exponentes, comprobar si un número de Mersenne es primo o no es algo que no puede hacerse mediante métodos "rudimentarios".

En 1774 Euler pudo probar que el número

$$M_{31} = 2\,147\,483\,647$$

es primo, y no fue hasta 1876 cuando el matemático francés Édouard Lucas se las arregló para demostrar que

$$M_{127} = 170\,141\,183\,460\,469\,231\,731\,687\,303\,715\,884\,105\,727$$

es un número primo. Durante 75 años fue el mayor primo conocido, y es el mayor número que se ha comprobado que es primo manualmente, sin la ayuda de un ordenador. En este libro expondremos el método empleado por Lucas para llevar a cabo su comprobación.

El desafío de Fermat En 1657 Fermat escribió una carta al vizconde William Brouncker, presidente de la *Royal Society,* en la que lamentaba la falta de interés por la aritmética entre los matemáticos de la época y, para ilustrar que ésta plantea problemas notables, propuso a los matemáticos ingleses este desafío:

> *Dado cualquier número que no sea un cuadrado, éste determina un número infinito de cuadrados tales que si el cuadrado se multiplica por el número dado y se le suma* 1 *al producto, el resultado es un cuadrado.*
>
> *Ejemplo: Sea* 3 *—que no es un cuadrado— el número dado. Cuando se multiplica por el cuadrado* 1 *y se le suma* 1*, el resultado es* 4*, que es un cuadrado.*
>
> *El mismo* 3 *multiplicado por el cuadrado* 16 *da un producto que, cuando se le suma* 1*, se convierte en* 49*, un cuadrado.*
>
> *Y es posible encontrar un número infinito de cuadrados, aparte de* 1 *y* 16*, que tienen la misma propiedad.*
>
> *Pero yo estoy pidiendo una regla general de solución para cualquier número no cuadrado que sea dado.*
>
> *Por ejemplo, se requiere encontrar un cuadrado tal que, si al producto del cuadrado y el número* 148 *o* 109 *o* 433 *se le suma* 1 *el resultado es un cuadrado.*

En términos modernos, Fermat preguntaba por un método para encontrar las soluciones enteras de la ecuación

$$x^2 - dy^2 = 1,$$

donde d no es un cuadrado perfecto.

Fermat no publicó el método con que era capaz de resolver estas ecuaciones, pero el hecho de que propusiera entre sus ejemplos casos como $d = 109$ demuestra que sabía resolverlas, pues ése es uno de los casos más difíciles.

La primera solución en respuesta a su desafío la publicó John Wallis, quien atribuyó la idea a Brouncker, si bien hay quien piensa que la solución era suya y que la atribución sólo fue una forma de ganarse su favor. El caso fue que Euler interpretó o recordó mal las palabras de Wallis y acabó atribuyendo la solución al matemático John Pell, al que Wallis citaba con frecuencia, pero no por esta cuestión en particular, y así hoy la ecuación se conoce como *ecuación de Pell.*

El Último Teorema de Fermat Pero el hecho por el que Fermat es más famoso se produjo poco después de su muerte, acaecida en 1665. Su hijo Samuel publicó varias notas inéditas de su padre, entre las que figuraba una nota que Fermat había escrito en el margen de su ejemplar de la *Aritmética* de Diofanto:[5]

> *Es imposible descomponer un cubo en dos cubos, un bicuadrado en dos bicuadrados, y en general, una potencia cualquiera, aparte del cuadrado, en dos potencias del mismo exponente. He encontrado una demostración realmente admirable, pero el margen de este libro es demasiado pequeño para albergarla.*

En otras palabras, Fermat afirmaba que la ecuación $x^n + y^n = z^n$ no tiene soluciones con las tres variables positivas cuando el exponente n es mayor que 2. Muchos matemáticos después de él trataron sin éxito de demostrar lo que pasó a conocerse como el *Último Teorema de Fermat,* y no fue hasta 1995 cuando el matemático británico Andrew Wiles publicó una demostración completa. No obstante, cabe señalar que el resultado probado por Wiles fue en realidad el último eslabón de una cadena que incluye muchos resultados precedentes, como el teorema de Ribet (publicado por Ken Ribet en 1990), que es el que conecta el trabajo de Wiles con el Último Teorema de Fermat.

Aquí veremos cómo resolver la ecuación de Pell y demostraremos todos los resultados que hemos mencionado hasta ahora, salvo el Último Teorema de Fermat, del que no se conoce ninguna demostración que no requiera potentísimos resultados algebraicos que trascienden con creces el contenido de este libro. Es prácticamente seguro que Fermat no demostró "su" teorema. Nunca afirmó haberlo hecho. Probablemente no tardaría en descubrir que la demostración que creía haber encontrado era incorrecta y no se molestó en tachar una nota en un margen de un libro de uso privado.

Lo que si que consta que demostró fue que la ecuación $x^4 + y^4 = z^4$ no tiene soluciones enteras no triviales, y esto sí que lo demostraremos en este libro. Más aún, veremos pruebas debidas a Gauss de que las ecuaciones $x^3 + y^3 = z^3$ y $x^5 + y^5 = z^5$ no tienen soluciones enteras no triviales.

La conjetura de Catalan En 1844, la *Journal für die reine und angewandte Mathematik,* más conocida como revista de Crelle, por su fundador, el matemático alemán August Leopold Crelle, publicaba esta nota:

Nota

> *Extraída de una carta dirigida al editor por el Sr. E. Catalan, profesor ayudante de la Escuela Politécnica de París.*
>
> *Le ruego, señor, tenga a bien enunciar en su revista el teorema siguiente, que yo creo que es verdadero, si bien todavía no he logrado demostrarlo completamente. Tal vez otros tengan más suerte:*

[5] No era la única nota. Por ejemplo, otra decía lo siguiente: *"¿Puede hallarse, entre los números enteros, un cuadrado distinto de* 25 *que, cuando se aumenta en* 2*, se vuelve un cubo? Esto podría parecer en principio difícil de analizar, pero puedo probar con una demostración rigurosa que* 25 *es el único cuadrado que es menor que un cubo en* 2 *unidades."* Resolveremos este problema en la página 199.

> *Dos números enteros consecutivos distintos de* 8 *y* 9 *no pueden ser potencias exactas. Dicho de otro modo, la ecuación* $x^m - y^n = 1$*, en la que las incógnitas son enteras y positivas, no admite más que una única solución.*

El autor de la carta (que no se tomó la molestia de puntualizar que los exponentes m y n tienen que ser mayores que 1) era el matemático belga Eugène August Catalan, que tenía entonces 30 años. Su "teorema" se conoce desde entonces como *Conjetura de Catalan,* y como tal permaneció hasta 2002, cuando fue demostrada por el matemático rumano Preda Mihăilescu.

Como en el caso del Último Teorema de Fermat, la demostración del que ya puede llamarse teorema de Mihăilescu es inabordable con las técnicas que presentaremos en este libro. No obstante, probaremos que se cumple cuando los exponentes son $m = 2$ o $n = 2$, así como un caso particular que se encuentra probado en *La armonía de los números,* un tratado escrito en 1343 por el rabino francés Levi ben Gershon, más conocido como Gersónides, quien demostró que las únicas potencias de 2 y 3 consecutivas son $(1, 2)$, $(2, 3)$, $(3, 4)$ y $(8, 9)$. En otras palabras, que la conjetura de Catalan es cierta cuando x, y toman los valores 2 y 3 (en cualquier orden).

Los resultados que hemos mencionado (con las excepciones indicadas) son sólo una pequeña muestra del tipo de problemas que pueden resolverse con las técnicas que presentaremos en este libro. Terminamos esta introducción con la presentación y un estudio preliminar de un problema famoso y muy antiguo.

El problema del ganado de Arquímedes En el canto XII de la Odisea, la hechicera Circe le hace esta profecía a Ulises:

> *Llegarás más tarde á la isla de Trinacria, donde pacen las muchas vacas y pingües ovejas del Sol. Siete son las vacadas, otras tantas las hermosas greyes de ovejas, y cada una está formada por cincuenta cabezas. Dicho ganado no se reproduce ni muere, y son sus pastoras dos deidades, dos ninfas de hermosas trenzas: Faetusa y Lampetia; las cuales concibió del Sol Hiperión la divina Neera. La veneranda madre, después que las dió a luz y las hubo criado, llevólas á la isla de Trinacria, allá muy lejos, para que guardaran las ovejas de su padre y las vacas de retorcidos cuernos. Si á éstas las dejares indemnes, ocupándote tan sólo en preparar tu regreso, aún llegaríais á Ítaca, después de pasar muchos trabajos; pero, si les causares daño, desde ahora te anuncio la perdición de la nave y la de tus amigos. Y aunque tú escapes, llegarás tarde y mal á la patria, después de perder a todos los compañeros.*

La profecía se cumplió. A pesar de las advertencias de Ulises, sus hombres saciaron su apetito sacrificando el ganado del Sol, y por ello Zeus destrozó su barco y sólo Ulises —que no había participado en el sacrilegio— sobrevivió.

La isla Trinacria de la Odisea (la isla triangular) se identificó con Sicilia, donde vivía Arquímedes. El escritor alemán Gotthold Ephraim Lessing trabajaba de bibliotecario en la Biblioteca del Duque Augusto, en Wolfenbüttel, donde un día encontró un manuscrito que contenía una carta que Arquímedes de Siracusa había dirigido a Eratóstenes de Cirene, que a la sazón enseñaba en Alejandría. La carta contenía un poema de 22 dísticos elegíacos en el que planteaba el problema del calcular la composición de los rebaños del Sol, que para Arquímedes no era la indicada en la Odisea. En 1773 Lessing publicó una traducción anotada con una solución incorrecta del problema. Una traducción libre del poema es la siguiente:

> Del Sol, ¡oh, extranjero!, los inmensos ganados,
> si de verdad eres sabio, cuenta con diligencia:
> Tantos fueron otrora, que los feraces prados
> cubría de Trinacria, —Sicilia— su presencia;
> Cuatro eran los rebaños de otros tantos colores:
> uno blanco de leche, negro otro, de apariencia
> de brillante azabache; de otro más, los mejores,
> rubios, melocotones el color emulaban;
> y el cuarto parecía salpicado de flores.
> En cada gran manada los toros superaban
> a cuantos jamás viste; así eran de imponentes,
> de abundantes, los blancos, que juntos superaban
> en la mitad y un tercio de los negros lucientes
> a los melocotón —recuérdalo, extranjero—,
> y también que los negros, no menos excelentes,
> un cuarto eran y un quinto de los del postrero
> de los divinos hatos, de salpicada tez,
> junto a todos los toros del rebaño tercero,
> color melocotón. Lejos de la escasez,
> de toros salpicados había, en proporción,
> como un sexto y un séptimo de blancos, y otra vez,
> cuantos machos formaban la grey melocotón.
> Así se repartían las vacas inmortales:
> Las blancas por sí solas —y aquí pon atención—
> un tercio eran y un cuarto de las reses totales
> de la grey distinguida por su negro color;
> Las negras resultaban ser en número iguales
> a un cuarto más un quinto del florido esplendor
> de aquellas salpicadas con sus toros reunidas.
> Y éstas, las salpicadas —de número inferior—

un quinto eran y un sexto de las reses incluidas
en el tercer rebaño, el de aterciopelada
tez de melocotón, cuyas vacas, distinguidas,
hasta un sexto y un séptimo de la blanca manada
en número alcanzaban. ¡Oh, extranjero! si contar
pudieras las cabezas —no en forma aproximada—
del ganado del Sol, llegando a precisar
cuántos toros robustos, cuántas vacas tenía
según cada color, no te habrán de llamar
lego ya en la aritmética. No obstante, aún faltaría
para que entre los sabios te vieras numerado;
Pero, vamos, prosigue, pues quedan todavía
otras dos condiciones sobre el sacro ganado:
Cuando los toros blancos a los negros se unían,
formaban la figura de un sólido cuadrado,
y tantos toros de alto cuantos de ancho cubrían
los campos de Trinacria. Sin embargo, cuando eran
los de melocotón los que a un tiempo pacían
junto a los salpicados, buscaban que los vieran
de tal forma dispuestos que, creciendo desde uno,
un triángulo inmenso parecía que fueran.
Si tu mente asimila como fuere oportuno
cuanto se ha dicho aquí, y atinas cada parte
del rebaño a medir, ¡Oh, extranjero! ninguno
tu triunfo negará, y piensa que al marcharte
lo harás glorificado, contado finalmente
entre los más expertos del aritmético arte,
por tu sabiduría, con creces excelente.

Si llamamos B, N, M, S al número de toros blancos, negros, melocotón y salpicados, respectivamente, y b, n, m, s al número correspondiente de vacas, las condiciones de la primera parte del enunciado son:

$$B = \left(\frac{1}{2} + \frac{1}{3}\right) N + M, \quad N = \left(\frac{1}{4} + \frac{1}{5}\right) S + M, \quad S = \left(\frac{1}{6} + \frac{1}{7}\right) B + M,$$

$$b = \left(\frac{1}{3} + \frac{1}{4}\right) (N + n), \quad n = \left(\frac{1}{4} + \frac{1}{5}\right) (S + s), \quad s = \left(\frac{1}{5} + \frac{1}{6}\right) (M + m),$$

$$m = \left(\frac{1}{6} + \frac{1}{7}\right) (B + b).$$

Consideremos las tres primeras:

$$B = \frac{5}{6} N + M, \quad N = \frac{9}{20} S + M, \quad S = \frac{13}{42} B + M.$$

Sustituyendo B en la tercera queda

$$B = \frac{5}{6} N + M, \quad N = \frac{9}{20} S + M, \quad S = \frac{65}{252} N + \frac{55}{42} M.$$

Es fácil convencerse de que este sistema de ecuaciones tiene exactamente las mismas soluciones que el original. Ahora sustituimos la expresión de N en la tercera ecuación, con lo que pasamos también a un sistema equivalente:

$$B = \frac{5}{6}N + M, \quad N = \frac{9}{20}S + M, \quad S = \frac{13}{112}S + \frac{395}{252}M.$$

Pero la tercera ecuación equivale a

$$S = \frac{1\,580}{891}M,$$

de donde concluimos que las soluciones de las tres ecuaciones son

$$B = \frac{742}{297}M, \quad N = \frac{178}{99}M, \quad S = \frac{1\,580}{891}M,$$

para cualquier valor de M. Ahora consideramos las cuatro ecuaciones siguientes:

$$b = \frac{7}{12}N + \frac{7}{12}n, \quad n = \frac{9}{20}S + \frac{9}{20}m, \quad s = \frac{11}{30}M + \frac{11}{30}m, \quad m = \frac{13}{42}B + \frac{13}{42}b.$$

Sustituyendo las expresiones que hemos obtenido para B, N, S queda

$$\begin{array}{rrrrcl}
b & -\frac{7}{12}n & & & = & \frac{623}{594}M \\
 & n & -\frac{9}{20}s & & = & \frac{79}{99}M \\
 & & s & -\frac{11}{30}m & = & \frac{11}{30}M \\
-\frac{13}{42}b & & & +m & = & \frac{689}{891}M
\end{array}$$

Ahora tenemos un sistema de ecuaciones lineales dependiente de un parámetro M. Vamos a resolverlo usando lo que se conoce como el *método de reducción de Gauss*. Si a la última ecuación le sumamos la primera multiplicada por 13/42 obtenemos un sistema equivalente:

$$\begin{array}{rrrrcl}
b & -\frac{7}{12}n & & & = & \frac{623}{594}M \\
 & n & -\frac{9}{20}s & & = & \frac{79}{99}M \\
 & & s & -\frac{11}{30}m & = & \frac{11}{30}M \\
 & -\frac{13}{72}n & & +m & = & \frac{3\,913}{3\,564}M
\end{array}$$

Ahora a la última le sumamos la segunda multiplicada por 13/72:

$$\begin{array}{rrrrcl}
b & -\frac{7}{12}n & & & = & \frac{623}{594}S \\
 & n & -\frac{9}{20}s & & = & \frac{79}{99}S \\
 & & s & -\frac{11}{30}m & = & \frac{11}{30}S \\
 & & -\frac{13}{160}s & +m & = & \frac{2\,951}{2\,376}S
\end{array}$$

Y finalmente a la última ecuación le sumamos la tercera multiplicada por 13/160:

$$\begin{array}{rrrrrr} b & -\frac{7}{12}n & & & = & \frac{623}{594}M \\ & n & -\frac{9}{20}s & & = & \frac{79}{99}M \\ & & s & -\frac{11}{30}m & = & \frac{11}{30}M \\ & & & \frac{4\,657}{4\,800}m & = & \frac{604\,357}{475\,300}M \end{array}$$

La última ecuación nos da ahora que

$$m = \frac{604\,357}{461\,043}M.$$

Sustituyendo en la tercera ecuación obtenemos s, lo que nos permite despejar n de la segunda y a su vez b de la primera. El resultado es:

$$B = \frac{742}{297}M, \quad N = \frac{178}{99}M, \quad M = \frac{1\,580}{891}M,$$

$$b = \frac{2\,402\,120}{1\,383\,129}M, \quad n = \frac{543\,694}{461\,043}M, \quad m = \frac{604\,357}{461\,043}M, \quad s = \frac{106\,540}{125\,739}M.$$

Así, para cada valor que demos a M, obtenemos una solución del sistema de siete ecuaciones. Hasta aquí el problema es un mero ejercicio de álgebra lineal. Ahora necesitamos un poco de aritmética, pues las soluciones al problema tienen que ser números enteros, para lo cual, la condición necesaria y suficiente es que M sea múltiplo del mínimo común múltiplo de los denominadores de las siete fracciones. Los denominadores son:

$$3^3 \cdot 11, \quad 3^2 \cdot 11, \quad 3^4 \cdot 11, \quad 3^3 \cdot 11 \cdot 4\,657, \quad 3^2 \cdot 11 \cdot 4\,657, \quad 3^3 \cdot 4\,657,$$

de donde el mínimo común múltiplo resulta ser $3^4 \cdot 11 \cdot 4\,657 = 4\,149\,387$. Las soluciones posibles son, entonces:

$$B = 10\,366\,482k, \quad N = 7\,460\,514k, \quad M = 4\,149\,387k, \quad S = 7\,358\,060k,$$

$$b = 7\,206\,360k, \quad n = 4\,893\,246k, \quad m = 5\,439\,213k, \quad s = 3\,515\,820k,$$

para cualquier número natural k. Con esto hemos conseguido que Arquímedes nos dé un aprobado, pero para llegar al sobresaliente necesitamos tener en cuenta las dos condiciones adicionales que se dan en la parte final del poema. La primera dice que $B + N$ es un cuadrado perfecto. Tenemos que

$$B + N = 17\,826\,996\,k = 2^2 \cdot 3 \cdot 11 \cdot 29 \cdot 4\,657\,k,$$

y ahora necesitamos de nuevo un poco de aritmética para concluir que, para que esta expresión sea un cuadrado perfecto, k tiene que ser de la forma

$$k = 3 \cdot 11 \cdot 29 \cdot 4\,657\,u^2 = 4\,456\,749\,u^2.$$

Esto obliga a que la solución sea de la forma:

$$\begin{array}{rlrl} B = & 46\,200\,808\,287\,018\,u^2 & b = & 32\,116\,937\,723\,640\,u^2 \\ N = & 33\,249\,638\,308\,986\,u^2 & n = & 21\,807\,969\,217\,254\,u^2 \\ M = & 18\,492\,776\,362\,863\,u^2 & m = & 24\,241\,207\,098\,537\,u^2 \\ S = & 32\,793\,026\,546\,940\,u^2 & s = & 15\,669\,127\,269\,180\,u^2 \end{array}$$

Hasta aquí no era difícil, pero falta incluir la última condición, que pide que

$$M + S = 51\,285\,802\,909\,803\,u^2$$

sea lo que los griegos llamaban un *número triangular,* es decir, un número de la forma

$$1 + 2 + 3 + 4 + \cdots + k,$$

para cierto k. Sin embargo, resolver el problema con esta última condición excede lo que podríamos considerar "aritmética elemental", así que tendremos que posponer la solución hasta la sección 10.5. El lector puede evaluar si sus conocimientos le bastan para acabar de resolverlo o si, por el contrario, necesitará estudiar algo más.

La ecuación de Pell, la ecuación de Catalan, o las ecuaciones que plasman el problema del ganado son ejemplos de *ecuaciones diofánticas,* es decir, de ecuaciones de las que se buscan sus soluciones enteras. No es un nombre muy afortunado, pues, en su *Aritmética,* los problemas que planteaba Diofanto consistían en encontrar soluciones racionales (no enteras) de diversas ecuaciones.

Terminamos aquí la "visita turística" a los contenidos de este libro y pasamos a exponerlos metódicamente, empezando desde la base más elemental.

Capítulo I

El álgebra de la escuela

Tal y como hemos indicado en el preámbulo, no vamos a suponer en el lector más conocimientos previos que cierta familiaridad con el álgebra y la aritmética básicas, lo suficiente para garantizar su competencia en la manipulación de expresiones algebraicas, despejar, operar, etc. No obstante, los conocimientos de un lector que pueda estar interesado en este libro pueden oscilar en una franja muy amplia, así que dedicamos este primer capítulo a discutir, precisar y organizar los preliminares algebraicos en los que nos apoyaremos en los capítulos siguientes (las propiedades básicas de los números naturales, enteros y racionales y de los polinomios). Más precisamente, el lector encontrará en este capítulo información que podemos clasificar como sigue:

1. Explicaciones más o menos detalladas sobre conceptos algebraicos elementales que el lector tal vez necesite o, por el contrario, pueda pasar por alto.

2. Demostraciones de hechos elementales que sin duda conocerá, pero que tal vez los aprendió en la escuela donde se los presentaron sin justificación alguna. Ante hechos de estas características, el lector puede optar por estudiar las demostraciones si las ignora o, por el contrario, considerar que no necesita que le demuestren aquello de cuya certeza no tiene duda alguna, y pasarlas también por alto.

3. Introducción de vocabulario algebraico adecuado para describir con precisión hechos que el lector conoce, aunque tal vez no los expresaría en tales términos (como que los números enteros forman un dominio íntegro, o que la relación de orden de los números enteros es compatible con la suma y con el producto).

4. Demostraciones de que algunas propiedades elementales que el lector conocerá sin duda sobre manipulación de expresiones numéricas son en realidad válidas en el contexto más general de determinadas estructuras algebraicas abstractas (como dominios íntegros, anillos ordenados, etc.), de modo que ser consciente de este grado de generalidad le permitirá desenvolverse con fluidez al operar con objetos algebraicos distintos de los "números clásicos"

(como, por ejemplo, los enteros de Gauss) sin más cuidado que aplicar únicamente en cada momento aquellas reglas familiares de cálculo que sean realmente válidas en el contexto considerado.

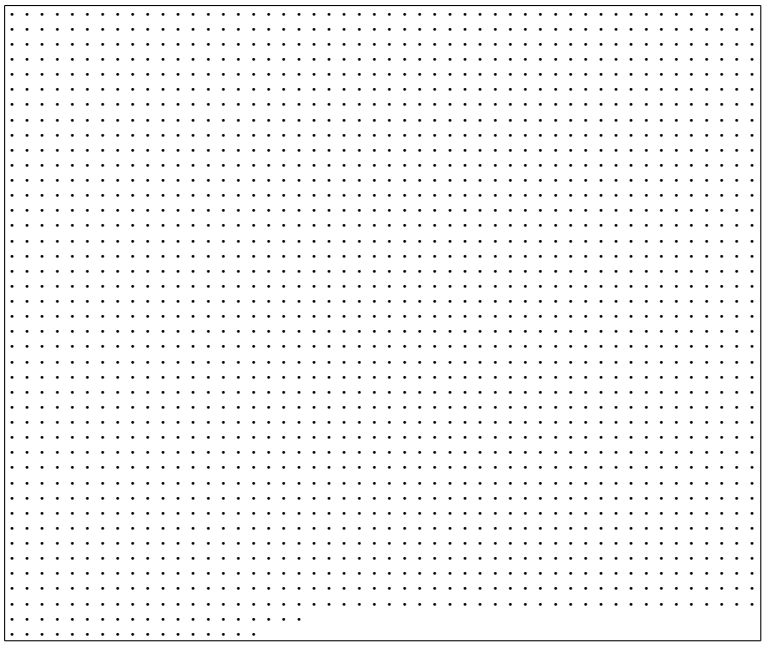
Así pues, el lector debería ir ojeando este capítulo y juzgar en cada momento si necesita o no leer con atención cada parte del mismo, en función de sus conocimientos previos y de sus intereses. Naturalmente, un lector con una buena base de álgebra abstracta podrá pasar directamente al capítulo siguiente.

1.1 Los números naturales

"Contar" fue sin duda el primer problema matemático al que se enfrentó la humanidad. Y la posibilidad de reflejar gráficamente de algún modo el resultado de un cómputo para referencia posterior surgió antes incluso de la aparición de la escritura. Cada cultura desarrolló su propia forma, más o menos afortunada, de representar gráficamente los *números naturales,* es decir, los resultados posibles de un cómputo. Nosotros usamos los numerales arábigos:

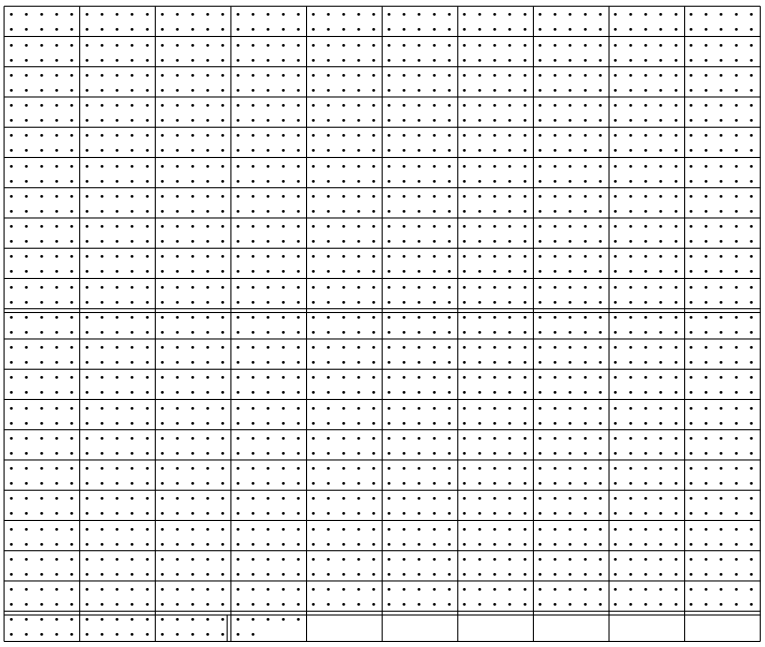
$$0, \quad 1, \quad 2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 7, \quad 8, \quad 9, \quad \ldots$$

que el lector conocerá sin duda, y que ha evolucionado a partir de los diversos sistemas de numeración que surgieron en la India entre los siglos I y IV de nuestra era. Aunque sería absurdo tratar de enseñar a contar aquí al lector, sí merece la pena detenernos a reflexionar sobre las características matemáticas de este sistema de numeración que lo hace superior a la mayor parte de los demás sistemas de numeración —si no a todos— que han surgido en la historia. Consideremos el problema de contar cuántos puntos hay aquí:



La característica principal nuestro sistema de numeración es que nos permite representar cualquier número natural sin usar más que los nombres específicos de los diez primeros y, más aún, nos permite contar cualquier conjunto de objetos sin necesidad de contar en ningún momento conjuntos de más de diez elementos.

En principio, si sólo sabemos contar hasta diez, ante la pregunta de cuántos puntos hay en el recuadro anterior sólo podremos responder que hay "muchos" (es decir, más de diez). Sin embargo, puestos a concretar, hay algo que sí que podemos hacer, y es agrupar los puntos en *decenas.* Eso sí que sabemos hacerlo, y el resultado es:



Si nos preguntan cuántas decenas hemos formado, la respuesta sigue siendo que "muchas" (más de diez), pero podemos observar que en el último grupo formado no hemos podido completar una decena, sino que nos han quedado 7 unidades sueltas.

En segundo lugar, aunque no sepamos cuántas decenas hay, lo que sí que podemos observar es que en cada fila de rectángulos hay diez decenas, es decir, una centena y, aunque el número de centenas sigue siendo "muchas", podemos observar que, descontando las 7 unidades sueltas, en la última fila queda un grupo de 3 decenas que no llegan para formar una centena.

En tercer lugar podemos agrupar las centenas en grupos de diez, es decir, en millares, y observamos que al hacerlo quedan 0 centenas sueltas que no lleguen a formar un millar y, más aún, esta vez los millares que obtenemos no son "muchos", sino que sabemos contarlos y resultan ser 2.

La conclusión es que el número de puntos dado consta de 2 millares, más 0 centenas sueltas, más 3 decenas sueltas más 7 unidades sueltas, y este número es el que representamos como 2 037.

Por supuesto, cuando contamos no lo hacemos agrupando decenas, centenas, etc. o, por lo menos, no necesitamos hacerlo, pues sabemos como prolongar arbitrariamente la sucesión

$$0,\quad 1,\quad 2,\quad 3,\quad 4,\quad 5,\quad 6,\quad 7,\quad 8,\quad 9,\quad 10,\quad 11,\quad 12,\quad 13,\quad \ldots$$

pero es muy relevante que, independientemente de cómo contemos, un número como 2 037 es "autoexplicativo", y significa precisamente lo que acabamos de recordar que significa.

Este sistema de numeración es *posicional* porque una misma cifra, como 2, puede significar dos unidades, o dos decenas, o dos centenas, etc., según su posición en el número. El hecho de que la base de numeración sea diez, es decir, que demos nombres individuales a los diez primeros números naturales, es anecdótico y no tiene más razón de ser que el hecho de que tenemos 10 dedos en las manos, por lo que 10 es el mayor número que podemos contar con los dedos.[1]

Por el contrario, otra característica de nuestro sistema de numeración que no tiene nada de anecdótica es la inclusión del 0 entre los números naturales, imprescindible[2] para que no haya ambigüedades a la hora de decidir si una cifra debe entenderse como un número de unidades, decenas, centenas, etc. La invención del 0 se atribuye al matemático y astrónomo indio Brahmagupta, en un trabajo aparecido en 628. Los árabes lo difundieron y hacia el siglo X empezó a ser usado en Europa. A finales de dicho siglo, Gerberto de Aurillac, el que más tarde se convertiría en el papa Silvestre II, difundió el uso de los números arábigos, y en 1 202 el matemático Leonardo Fibonacci (también conocido como Leonardo de Pisa) escribió el *Liber abaci,* en el que presenta el sistema de numeración arábigo junto con diversos problemas que se engloban en lo que hoy conocemos como teoría de números. Sin embargo, no fue hasta el siglo XV cuando su uso se generalizó en Europa, especialmente a partir de la invención de la imprenta.

Antes de la implantación del sistema decimal, en Europa se usaba la numeración romana, que, además de que no proporciona un criterio simple y homogéneo para prolongar indefinidamente la sucesión de numerales:

$$\text{I},\quad \text{II},\quad \text{III},\quad \text{IV},\quad \text{V},\quad \text{VI},\quad \text{VII},\quad \text{VIII},\quad \text{IX},\quad \text{X},\quad \ldots$$

era muy poco adecuada para operar. Para cualquier cálculo mínimamente sofisticado el uso del ábaco era poco menos que imprescindible. Sin embargo, uno de los propósitos del *Liber abaci* era mostrar cómo el sistema de numeración

[1] Esto es cuestión de ingenio, pues los antiguos babilonios usaban el pulgar para señalar las 12 falanges de los cuatro dedos restantes de la mano, y así podían contar hasta 12 con una mano, y luego usaban los cinco dedos de la otra para contar docenas, con lo que en total podían contar hasta 60 con los dedos. Esto hizo que desarrollaran un sistema de numeración posicional de base 60, pero, como usar 60 cifras sería farragoso, nombraban cada cifra del 1 al 59 con otro sistema de numeración posicional, pero esta vez en base 10.

[2] Los babilonios usaban un sistema de numeración posicional sin que el 0 tuviera un nombre asignado, dejando huecos entre las cifras, pero esto hacía inevitable que el significado de algunos números tuviera que interpretarse según el contexto, pues, por ejemplo, nada distinguía la forma de escribir 3 o $3 \cdot 60 = 180$.

decimal volvía innecesario el ábaco para hacer operaciones, pues la representación decimal equivale a la representación de los números en el ábaco, y permite realizar las operaciones aritméticas a través de algoritmos sencillos.[3]

Damos por hecho que el lector está familiarizado con la numeración arábiga, lo cual —además de entender cómo la representación de cada número describe con precisión la cantidad a la que hace referencia— supone ser capaz de generar y prolongar indefinidamente la sucesión de los números naturales, y entender cómo se usa para contar conjuntos de objetos y concluir, por ejemplo, que las arañas tienen 8 patas, o que los icosaedros tienen 30 aristas.

El orden de los números naturales Es difícil concebir que alguien conozca los números naturales sin conocer también su ordenación:

> Si m y n son dos números naturales, decimos que m *es menor que* n ($m < n$) o que n *es mayor que* m ($n > m$) si m aparece antes que n en la sucesión de los números naturales o, en términos del cómputo, que en un conjunto de n cosas hay suficientes para extraer m de ellas (sin agotarlas todas).

Si queremos evitar la precisión final y no excluir la posibilidad de que m sea igual a n ($m = n$) escribimos $m \leq n$ (o $n \geq m$) y leemos *"m es menor o igual que n"* o *"n es mayor o igual que m"*.

El resumen 1.1 recoge las propiedades básicas de la ordenación de los números naturales, que vamos a comentar a continuación. Ante todo, conviene aislar las más básicas de todas para definir un concepto general de "conjunto ordenado".

En general, ordenar los elementos de un conjunto supone fijar un criterio que determine cuándo un objeto es menor o igual que otro ($m \leq n$), pero no todo criterio es aceptable. Por ejemplo, si un criterio estableciera que unos objetos cumplen $m \leq n$ y $n \leq r$, pero al mismo tiempo negara que $m \leq r$, no podría interpretarse como una relación de orden.[4]

Diremos que una relación $m \leq n$ entre los elementos de un conjunto es una *relación de orden (total)* si cumple las cuatro propiedades básicas consignadas en el resumen 1.1. Esto lo cumple claramente cualquier criterio que pueda interpretarse como que dispone los elementos del conjunto uno detrás de otro, o uno antes que otro, o uno a la izquierda de otro. En tal caso decimos también que

[3] Hay quienes niegan al 0 su condición de número natural —incluso expertos en teoría de números—, pero ése es un lujo que pueden permitirse porque nunca trabajan exclusivamente con números naturales, sino que siempre consideran como mínimo números enteros, entre los cuales incluyen al 0. Dicen que "no es natural" contar conjuntos con 0 cosas, sin tener en cuenta que el 0 es indispensable hasta para nombrar los propios números naturales con el sistema que empleamos actualmente, basado en contar "idealmente" grupos de unidades, decenas, centenas, etc., de modo que no es raro que en la descomposición decimal de un número nos encontremos con que contiene 0 decenas o 0 centenas, etc.

[4] Por ejemplo, puede ocurrir que m sea amigo de n y que n sea amigo de r y que al mismo tiempo m no sea amigo de r. Esto significa que, aun admitiendo que la amistad entre un conjunto de personas pueda ser un criterio objetivamente determinado, no es un criterio de ordenación.

Resumen 1.1: **Propiedades generales de las relaciones de orden**

Propiedad reflexiva:	$m \leq m$
Propiedad antisimétrica:	Si $m \leq n$ y $n \leq m$, entonces $m = n$
Propiedad transitiva:	Si $m \leq n$ y $n \leq r$, entonces $m \leq r$
Dicotomía:	O bien $m \leq n$ o bien $n \leq m$.

Propiedades específicas del orden de los números naturales

- *Todo número natural tiene un siguiente (un sucesor inmediato).*
- *Todo número natural distinto de* 0 *tiene un anterior inmediato.*
- **Principio de buena ordenación** *Todo conjunto no vacío de números naturales tiene un mínimo elemento.*
- *Todo conjunto no vacío de números naturales acotado superiormente tiene un máximo elemento.*

el conjunto sobre el que está definida dicha relación es un conjunto (totalmente) ordenado[5] y podemos considerar también en él la relación estricta $m < n$ que resulta de añadir a $m \leq n$ la exigencia de que $m \neq n$.

La ordenación de los números naturales cumple obviamente estas propiedades básicas, pero además cumple otras específicas que la diferencian de las ordenaciones de otros conjuntos. El resumen 1.1 contiene también estas propiedades específicas básicas. Todas ellas son evidentes.

La primera afirma que todo número natural tiene un siguiente, es decir, un mínimo número natural posterior a él, y la segunda que todo número natural distinto de 0 tiene un anterior (inmediato), es decir un máximo número natural anterior a él. Notemos que es lo mismo decir que n es el siguiente de m o que m es el anterior (inmediato) de n.

Estas dos propiedades expresan la esencia de lo que es la sucesión de los números naturales, que consta del 0, del siguiente del 0, del siguiente del siguiente del 0, y de todos los números que van generándose mediante este proceso (y ninguno más).

> En términos prácticos, si $m < n$ y n es, concretamente, el siguiente de m, esto se traduce en que $m < r$ es equivalente a $n \leq r$, pues n es el menor de los números posteriores a m y r es uno de ellos.
>
> Similarmente, $r < n$ es equivalente a $r \leq m$, pues en caso contrario tendría que ser $m < r$, lo que a su vez implica que $n \leq r$.

El principio de buena ordenación afirma que todo conjunto no vacío de números naturales tiene un mínimo elemento, es decir, uno que es menor que los

[5] Si no exigimos que la relación cumpla la propiedad de dicotomía tenemos lo que se conoce como un orden parcial y un conjunto parcialmente ordenado.

demás elementos del conjunto. Esto es inmediato: si vamos generando la sucesión de los números naturales, en algún momento aparecerá un elemento del conjunto considerado, y el primero que aparezca es el mínimo del conjunto.

En particular, el conjunto formado por todos los números naturales tiene un mínimo elemento, que es, por supuesto, el 0.

Notemos cómo se usa en la práctica el principio de buena ordenación:

> *Siempre que hayamos justificado la existencia de un número natural con una determinada propiedad, estaremos legitimados a considerar, si nos interesa, el menor número natural con dicha propiedad.*

No es cierto, en cambio, que todo conjunto no vacío de números naturales tenga un máximo elemento. Eso es precisamente lo que distingue a los conjuntos finitos de los infinitos. Fijado un conjunto no vacío de números naturales, podemos ir recorriendo la sucesión entera de los números naturales contando los elementos del conjunto a medida que van apareciendo. Pero pueden darse dos casos: o bien llegamos a un elemento del conjunto tras el cual ya no haya ninguno más, en cuyo caso el conjunto es *finito,* hemos calculado cuántos elementos tiene y el último elemento que ha aparecido es su máximo, o bien puede suceder que nunca dejen de aparecer nuevos elementos del conjunto, en cuyo caso el proceso de cómputo no termina nunca, el conjunto es *infinito* y no tiene máximo elemento.[6]

Ahora bien, si podemos asegurar que el conjunto considerado tiene una *cota superior* c, es decir, que todos los elementos del conjunto cumplen $r \leq c$, entonces podemos asegurar que el conjunto es finito y tiene un máximo elemento, que será el último elemento del conjunto que haya aparecido al generar la sucesión de los números naturales hasta c.

Suma de números naturales La humanidad conoce el concepto de suma desde el neolítico:

> *Si tenemos un conjunto con m cosas y otro con n cosas distintas de las anteriores, entonces $m + n$ es el número de cosas que tiene el conjunto que resulta de reunir los dos conjuntos dados en uno solo.*

De aquí se desprenden una serie de propiedades elementales recogidas en el resumen 1.2 y que conviene discutir brevemente para apreciar su relevancia:

En primer lugar tenemos la propiedad asociativa. Si tenemos un conjunto con m cosas, otro con n cosas distintas y otro con r cosas más, distintas de las

[6] Notemos que este proceso de enumeración que hemos descrito es meramente ideal, pues no tenemos garantía de que podamos realizarlo en la práctica. Por un lado, podría suceder que no tuviéramos medios de saber si un número natural dado forma parte o no del conjunto considerado —con lo que no sabríamos si contarlo o no— y, por otro lado, podría suceder que el conjunto fuera finito y ya hubiéramos contado todos sus elementos, pero no supiéramos si todavía quedan más por aparecer.

Resumen 1.2: Propiedades de la suma de números naturales

Asociativa:	$(m+n)+r = m+(n+r)$
Conmutativa:	$m+n = n+m$
Elemento neutro:	$n+0 = n$
Simplificación:	Si $m+r = n+r$, entonces $m = n$
Compatibilidad con el orden:	Si $m \leq n$, entonces $m+r \leq n+r$

El siguiente de n es $n+1$.

$m \leq n$ si y sólo si existe un r tal que $m+r = n$.

anteriores, entonces $(m+n)+r$ es el número de cosas que tenemos al reunir los dos primeros montones en uno solo, y luego añadir las cosas del tercero, mientras que $m+(n+r)$ es el número de cosas que resulta de unir primero los dos últimos y luego añadir el primero. Evidentemente, en ambos casos se trata del número total de cosas que contienen los tres montones juntos, luego ambos resultados deben coincidir. La relevancia de esta trivialidad es que nos permite escribir expresiones como

$$5+12+7+4+1+5$$

sin necesidad de especificar si nos referimos a

$$5+((12+7)+(4+(1+5)))$$

o bien a

$$(((5+12)+7)+(4+1))+5$$

o a cualquier otra agrupación de los sumandos. La propiedad asociativa garantiza que cualquiera de ellas dará el mismo resultado.

Los matemáticos estudian muchas operaciones diferentes de la suma de números naturales, pero muy raras veces se prestan a estudiar operaciones que no sean asociativas, precisamente porque la falta de asociatividad complica terriblemente el análisis de las expresiones que se generan al aplicar repetidamente la operación.

La propiedad conmutativa se justifica análogamente a la asociativa, y simplifica aún más la manipulación de sumas sucesivas, pues nos permite hacer transformaciones del estilo de

$$1+2+3+4+5+6+7+8+9+10 = 1+10+2+9+3+8+4+7+5+6$$

$$= 11+11+11+11+11+11.$$

No es infrecuente que los matemáticos trabajen con operaciones no conmutativas, y es fundamental no realizar manipulaciones que presuponen la conmutatividad cuando se trabaja con operaciones que carecen de esta propiedad.

Dejamos que el lector se convenza por sí mismo de la obviedad de las propiedades siguientes contenidas en el resumen 1.2 y pasamos a comentar las dos últimas, que relacionan la suma con la relación de orden. La primera de ellas afirma que el siguiente de un número natural n, es decir, el menor número mayor que n, el número que viene justo después de n en la sucesión de los números naturales, puede describirse más fácilmente como $n+1$.

> En la práctica, esto justifica manipulaciones como que $n < r$ es equivalente a $n+1 \leq r$, o que $n \leq r$ es equivalente $n < r+1$.

Otra consecuencia es que todo número $n \geq 1$ puede expresarse como

$$n = \overbrace{1+\cdots+1}^{n \text{ veces}}.$$

La última propiedad muestra que la relación de orden está determinada por la suma, pues $m \leq n$ es equivalente a que n puede obtenerse de m sumándole un cierto número natural r. Esto nos da una interpretación alternativa de la suma de números naturales:

$$m + r = m + \overbrace{1+\cdots+1}^{r \text{ veces}}.$$

Una suma $m+r$ es el número que resulta de sumarle r veces 1 a m o, equivalentemente, de aplicarle r veces la operación "paso al siguiente" (lo cual vale también trivialmente si $r = 0$).

> Otra consecuencia es que el resultado de una suma es siempre mayor o igual que cada uno de los sumandos, y en particular vemos que si $m + n = 0$, necesariamente $m = n = 0$.

Para completar la revisión de las propiedades elementales de la suma de números naturales es obligado mencionar el algoritmo que el lector conoce sin duda para sumar números a partir de su desarrollo decimal:

$$\begin{array}{r} 87 \\ +\,56 \\ \hline 143 \end{array}$$

No vamos a describirlo con detalle porque el lector lo habrá aprendido sin duda en la escuela, aunque tal vez el lector quiera reflexionar para convencerse de que está justificado, es decir, que el proceso que conoce para sumar números naturales realmente proporciona la suma de los números a los que se aplica.

La existencia de calculadoras y ordenadores hace que estos algoritmos aritméticos sean ahora de escasa utilidad, pero ya hemos señalado que en su momento fueron uno de los factores decisivos para la consolidación del sistema de numeración posicional.

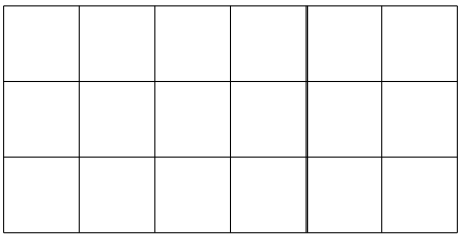
Resumen 1.3: Propiedades del producto de números naturales

Asociativa:	$(mn)r = m(nr)$
Conmutativa:	$mn = nm$
Distributiva:	$m(n + r) = mn + nr$
Elemento neutro:	$n \cdot 1 = n$

Multiplicación de números naturales El concepto de multiplicación de números naturales es tan básico y conocido como el de la suma:

> *El producto $m \cdot n$ de dos números naturales es el resultado de sumar n sumandos iguales a m.*

Aquí hay que entender que, por definición, $m \cdot 0 = 0$. El resumen 1.3 recoge las propiedades básicas de la multiplicación. La justificación más clara de dichas propiedades es geométrica. Por ejemplo, para justificar la propiedad conmutativa podemos pensar en un rectángulo formado por cuadrados:



El número total de cuadrados puede calcularse igualmente como

$$3 + 3 + 3 + 3 + 3 + 3 = 6 + 6 + 6,$$

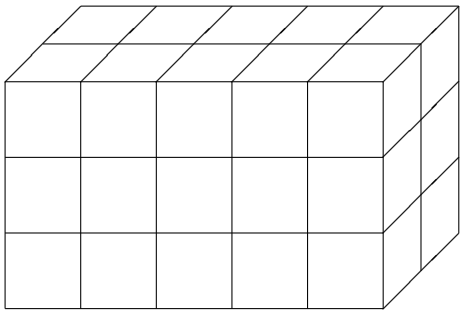
lo que muestra que $3 \cdot 6 = 6 \cdot 3$, y es claro que el argumento vale para factores arbitrarios (no nulos, si se quiere, pero si algún factor es nulo es claro que $n \cdot 0 = 0 \cdot n = 0$).

Si descomponemos $6 = 4 + 2$, la misma figura muestra por qué

$$3 \cdot (4 + 2) = 3 \cdot 4 + 3 \cdot 2$$

así como que el argumento vale para números arbitrarios (tratando aparte, si se quiere, el caso en que alguno de ellos sea nulo). Esto justifica la propiedad distributiva.

Para justificar la propiedad asociativa podemos pensar en un prisma formado por cubos apilados. El número total de cubos en la figura siguiente puede calcularse como el número de cubos que forman la base $(5 \cdot 2)$ multiplicado por las tres capas de cubos: $(5 \cdot 2) \cdot 3$, pero también como el número de cubos que forman una cara lateral $(2 \cdot 3)$ multiplicado por las cinco rebanadas de cubos: $5 \cdot (2 \cdot 3)$, y esto prueba que $(5 \cdot 2) \cdot 3 = 5 \cdot (2 \cdot 3)$.



De nuevo es claro que el argumento se puede aplicar a factores no nulos arbitrarios, y el caso en que algún factor es nulo es trivial porque el producto da 0 en cualquier caso.

Tal y como hemos señalado al comentar la asociatividad de la suma, la propiedad asociativa del producto justifica que escribamos expresiones como

$$5 \cdot 12 \cdot 534 \cdot 5 \cdot 2$$

sin necesidad de especificar cómo tienen que agruparse los factores en el cálculo.

No hemos incluido propiedades que relacionen el producto con la relación de orden porque todas ellas se deducen fácilmente de las enunciadas y de las propiedades de la suma. Por ejemplo:

1. *Si $m \leq n$, entonces $mr \leq nr$.*

   En efecto, basta tener en cuenta que existe un s tal que $m + s = n$, con lo que $mr + sr = (m + s)r = nr$, y esto equivale a que $mr \leq nr$.

2. *Si $mn = 0$, entonces $m = 0$ o $n = 0$.*

   En efecto, si $m \neq 0$, entonces $1 \leq m$, luego $n = 1 \cdot n \leq mn = 0$, luego $n = 0$.

3. *Si $mr = nr$ y $r \neq 0$, entonces $m = n$.*

   En efecto, no perdemos generalidad si suponemos que $m \leq n$. Entonces existe un s tal que $m + s = n$, luego $mr + sr = nr = nr + 0$, luego $sr = 0$, luego $s = 0$, luego $m = n$.

Para terminar mencionamos que el lector conocerá sin duda el algoritmo que permite multiplicar dos números a partir de su expresión decimal:

$$\begin{array}{r} 4120 \\ \times\ \ 164 \\ \hline 16480 \\ 24720 \\ +4120 \\ \hline 675680 \end{array}$$

Nuevamente, el lector podría meditar sobre que este proceso, que probablemente le enseñaron de niño sin justificación alguna, realmente proporciona el producto de los números a los que se aplica.

Resta de números naturales La suma y la multiplicación tienen la peculiaridad de que se pueden calcular para cualquier par de números naturales dados. No ocurre lo mismo con la resta:

> *La resta de dos números naturales m y n es el único número natural $x = m - n$ que cumple $n + x = m$.*

En otras palabras, $m - n$ son las cosas que quedan cuando de un conjunto de m cosas quitamos n de ellas. Es obvio que la condición necesaria y suficiente para que esté definida la resta $m - n$ de dos números naturales es que $n \leq m$.

No merece la pena enunciar propiedades sobre la resta, porque en cuanto introduzcamos los números enteros en la sección siguiente dichas propiedades serán casos particulares de las de la suma. Mencionamos, no obstante, la existencia del conocido algoritmo para restar números naturales a partir de su expresión decimal:

$$\begin{array}{r} 412 \\ -\ 58 \\ \hline 354 \end{array}$$

División de números naturales Podemos ver la resta $m-n$ de dos números naturales como la solución (si es que existe) de la ecuación $n + x = m$, e igualmente podemos plantearlos el cálculo de la solución (si es que existe) de la ecuación $mx = n$.

> Se dice que un número natural m es *divisor* de otro n (o que n es *múltiplo* de m o *divisible* entre m) y se representa por $m \mid n$, si existe un número x tal que $mx = n$. En tal caso, si $m \neq 0$, dicho x está unívocamente determinado, se llama *cociente* de n entre m y se representa por $x = n/m$.

Notemos que, en efecto, si $m \neq 0$ y se cumple $mx = n = my$, entonces podemos simplificar m para concluir que $x = y$, luego el cociente está unívocamente determinado (en caso de existir). Por otro lado, si $m = 0$ sólo existe un cociente cuando $n = 0$, y en tal caso todos los números x cumplen $0 \cdot x = 0$, por lo que no podemos asignar un valor a la expresión $0/0$.

Usaremos la notación $m \nmid n$ para indicar que m no divide a n.

El resumen 1.4 contiene algunas propiedades elementales de la relación de divisibilidad que se demuestran fácilmente a partir de la definición.[7] De las dos últimas propiedades se deduce, más en general, que si tenemos una relación

$$m + r = n,$$

entonces un número s divide a dos de los términos si y sólo si divide al tercero.

[7] Las tres primeras son la propiedad reflexiva, antisimétrica y transitiva, que muestran que la divisibilidad de números naturales es una relación de orden parcial, es decir, una relación de orden que no cumple el principio de dicotomía, pues, por ejemplo, $2 \nmid 3$ y $3 \nmid 2$.

Resumen 1.4: Propiedades de la divisibilidad de números naturales

- $m \mid m$
- Si $m \mid n$ y $n \mid m$, entonces $m = n$.
- Si $m \mid n$ y $n \mid r$, entonces $m \mid r$.
- Si $m \mid n$ y $n \neq 0$, entonces $m \leq n$.
- Si $m \mid n$ y $m \mid r$, entonces $m \mid n + r$.
- Si $m \mid n$, $m \mid r$ y $n \leq r$, entonces $m \mid r - n$.

Nuevamente no merece la pena enunciar propiedades de los cocientes de números naturales pues éstas se enunciarán mucho más cómodamente en cuanto introduzcamos los números racionales.

Un resultado fundamental en la teoría de números es que es posible considerar un concepto más amplio de división entre números naturales que puede realizarse (casi) con números cualesquiera. Se conoce como *división euclídea:*

> *Dados dos números naturales D (dividendo) y $d \neq 0$ (divisor), existen unos únicos números naturales c (cociente) y r (resto) tales que*
>
> $$D = dc + r, \qquad r < d.$$

El resultado es obvio: si tomamos un conjunto con D cosas, podemos ir formando con ellas grupos de d cosas hasta que ya no sea posible formar más grupos. Si nos han salido c grupos, nos habrán sobrado $r = D - dc$ cosas, y este número tiene que ser menor que d, pues de lo contrario habríamos podido formar un grupo más.

La unicidad tampoco ofrece dudas: si tratamos de plantear una igualdad similar con un cociente $c' < c$, entonces nos sobrarán más de d cosas, y es imposible obtener una igualdad similar con $c' > c$ porque c es el máximo número de grupos de d cosas que podemos formar con las D cosas dadas. Por lo tanto el valor del cociente c está unívocamente determinado, y eso hace que el resto $r = D - dc$ también sea único.

El lector conocerá sin duda un algoritmo para calcular divisiones euclídeas a partir de las expresiones decimales de los números dados:

$$\begin{array}{r|l} 3265 & 7 \\ \cline{2-2} 46 & 466 \\ 45 & \\ 3 & \end{array}$$

Este algoritmo permite determinar si un número dado d no nulo divide o no a otro D, pues $d \mid D$ equivale a que el resto de la división euclídea sea $r = 0$.

Bases de numeración Como es habitual, estamos representando los números naturales en base 10, es decir, que usamos 10 cifras para representar los números del 0 al 9 y en función de ellas expresamos todos los demás. Pero —como ya hemos señalado— la única razón de fondo para esto es que tenemos 10 dedos en las manos. En algunos contextos es conveniente usar otras bases de numeración. Por ejemplo, en informática es esencial considerar representaciones en base 2 —que son lo único que, en última instancia, entienden los ordenadores— y, como los números en base 2 son muy largos, resulta útil abreviarlos en base 16 y a veces en base 8, cuya relación con la base 2 es más directa y sencilla.

Por ejemplo, en la representación *hexadecimal* de un número natural se usan las cifras

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

de modo que, por ejemplo:

$$\mathrm{B3FC}_{16} = 11 \cdot 16^3 + 3 \cdot 16^2 + 15 \cdot 16 + 12 = 46\,076_{10}$$

Para pasar de base 10 a otra base, por ejemplo, 16, basta realizar divisiones sucesivas. Por ejemplo:

$$\begin{array}{r|l} 46076 & 16 \\ \cline{2-2} 140 & 2879 \;\big|\; 16 \\ \cline{3-3} 12 & 127 \quad 179 \;\big|\; 16 \\ \cline{4-4} & 159 \quad 19 \quad 11 \\ & 15 \quad 3 \end{array}$$

La primera división nos dice que el número consta de 2 879 "decenas hexadecimales" y que sobran 12 = C unidades, luego su menor cifra hexadecimal es C.

La segunda división nos indica que las 2 879 "decenas hexadecimales" contienen 179 "centenas hexadecimales" y sobran 15 = F "decenas hexadecimales", por lo que la segunda cifra del número es F.

La tercera división nos indica que las 179 "centenas hexadecimales" contienen 11 = *B* "millares hexadecimales" y que sobran 3 "centenas hexadecimales", por lo que la tercera cifra hexadecimal es 3 y la cuarta es B.

Realizando divisiones sucesivas entre 2 podríamos obtener la representación del mismo número en base 2, pero hay una alternativa más simple para pasar de base 16 a base 2. Basta reemplazar cada cifra hexadecimal por su representación binaria de cuatro dígitos:

0	0000	4	0100	8	1000	C	1100
1	0001	5	0101	9	1001	D	1101
2	0010	6	0110	A	1010	E	1110
3	0011	7	0111	B	1011	F	1111

Así:

$$\mathrm{B3FC}_{16} = 1011\,0011\,1111\,1100_2.$$

Los mismos algoritmos que usamos para operar en base 10 pueden usarse sin cambio alguno para operar en cualquier otra base.

1.2 Los números enteros

Los números naturales están vinculados a la noción de cómputo y, pensando en términos del cómputo, es obvio que el número de elementos de un conjunto no puede ser inferior a 0 y que de un conjunto de 5 elementos no podemos extraer 7, por lo que la resta $5 - 7$ no tiene sentido. Sin embargo, los números pueden usarse para representar otros conceptos distintos de la cantidad de elementos, y en los que las cantidades inferiores a 0 sí que tienen sentido. Por ejemplo, si asignamos un número a cada planta de un edificio, de modo que el 0 corresponde a la planta baja, ya es cuestionable que, estando en el piso 5, no podamos descender 7 plantas. Será posible si el edificio tiene al menos dos niveles de sótano por debajo de la planta baja.

Así pues, si pretendemos usar números para representar las plantas de un edificio, es razonable asociar el 0 a la planta baja, usar números positivos para los pisos situados sobre ella, pero también números negativos o "bajo cero" para las plantas subterráneas. Del mismo modo que el proceso de cómputo lleva al concepto abstracto de número natural, contextos como el que acabamos de analizar llevan a abstraer los llamados números enteros:

$$\cdots \quad -5, \quad -4, \quad -3, \quad -2, \quad -1, \quad 0, \quad +1, \quad +2, \quad +3, \quad +4, \quad +5, \quad \cdots$$

> Un *número entero* m es una expresión de la forma $+n$ o $-n$, donde n es un número natural al que se llama *valor absoluto* de m y se representa por $|m|$, con el convenio de que dos números enteros son iguales si y sólo si tienen el mismo signo y el mismo valor absoluto, excepto si el valor absoluto es 0, en cuyo caso convenimos que $+0 = -0$. Un número entero con valor absoluto no nulo es *positivo* o *negativo* según si su signo es $+$ o $-$.

Identificaremos los números naturales

$$0, \quad 1, \quad 2, \quad 3, \quad 4, \quad \ldots$$

con los números enteros

$$0, \quad +1, \quad +2, \quad +3, \quad +4, \quad \ldots$$

de modo que podemos pensar que los números enteros se obtienen de añadir los números negativos[8]

$$\ldots \quad -4, \quad -3, \quad -2, \quad -1$$

a los números naturales que ya teníamos.

[8] Es curioso que las matemáticas europeas tardaron bastante tiempo en incorporar los números negativos. Eran desconocidos para los griegos. Por ejemplo, en la *Aritmética* de Diofanto la ecuación $4x + 20 = 4$ es calificada de "absurda". La primera referencia histórica a los números negativos aparece en los *Nueve capítulos del arte matemático* de Jiu zhang suanshu, redactado entre el siglo II a.C. y el siglo II d.C., aunque puede haberse basado en material más antiguo. El uso en Europa procede, como la numeración decimal, de la matemática india. Brahmagupta hablaba de "fortunas" y "deudas" para referirse a los números enteros. El primer tratamiento moderno de los números enteros en Europa aparece en la *Ars Magna* de Gerolamo Cardano, publicada en 1545.

El orden de los números enteros Definimos como sigue la ordenación de los números enteros:

> *Se dice que un número entero m es menor o igual que otro n (y lo expresamos mediante $m \leq n$) si se da uno de los casos siguientes:*
>
> 1. *Ambos son positivos y $|m| \leq |n|$.*
> 2. *Ambos son negativos y $|m| \geq |n|$.*
> 3. *m es negativo o 0 y n es positivo o 0.*

Es claro que esto equivale a que m está situado antes (a la izquierda) de n en la disposición de los enteros que hemos mostrado al principio de esta sección,[9] y de aquí es inmediato que la relación que acabamos de definir satisface las cuatro propiedades que definen una relación de orden recogidas en el Resumen 1.1.

Notemos en particular que si m y n son números naturales, entonces se cumple $m \leq n$ en el sentido que ya teníamos definido para números naturales si y sólo si se cumple en el sentido que acabamos de introducir para números enteros. Podemos expresar esto diciendo que la relación de orden de los números enteros extiende a la que ya teníamos sobre los números naturales.

En términos de la relación de orden, se cumple que los números enteros positivos son los mayores que 0 y los negativos los menores que 0.

También es claro que todo número entero tiene un inmediato anterior y un inmediato posterior.

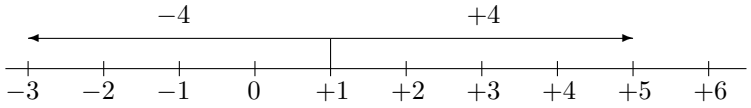
Los números enteros no cumplen el principio de buena ordenación, pero es obvio que cumplen la versión débil siguiente:

> *Si un conjunto de números enteros no vacío está acotado superiormente, tiene un máximo elemento, y si está acotado inferiormente, tiene un mínimo elemento.*

Aquí hay que entender que un conjunto está acotado superiormente (inferiormente) si tiene una *cota superior (inferior)* c, es decir, un número tal que todo elemento del conjunto cumple $r \leq c$ ($c \leq r$).

Suma de números enteros La suma de números enteros tiene una interpretación muy clara si pensamos en los enteros como indicadores de la posición en un edificio: si un número entero m corresponde a una determinada planta del edificio y le sumamos un número positivo $+n$, el resultado es el piso al que se llega ascendiendo n plantas desde la planta m, mientras que si sumamos un número negativo $-n$ el resultado es el piso al que se llega descendiendo n plantas desde la planta m (mientras que sumar 0 supone no moverse).

[9] Si pensamos que los números enteros se corresponden con los pisos de un edificio con infinitas plantas por encima de la planta baja e infinitos niveles de sótanos, entonces $m \leq n$ se interpreta como que la planta m está por debajo de la planta n.



La figura ilustra las sumas $(+1)+(+4)=+5$ y $(+1)+(-4)=-3$. Esto nos lleva a la definición siguiente:

> *La suma $m+n$ de dos números enteros m y n se define según los casos siguientes:*
>
> 1. *Si ambos sumandos son positivos o 0, la suma es el número positivo o 0 de valor absoluto $|m|+|n|=m+n$.*
> 2. *Si ambos sumandos son negativos o 0, la suma es el número negativo o 0 de valor absoluto $|m|+|n|$.*
> 3. *Si un sumando es positivo y otro negativo, la suma es el número cuyo valor absoluto es la resta de los valores absolutos de los sumandos (el mayor menos el menor) con el signo del sumando con mayor valor absoluto.*

Lo primero que conviene destacar de esta definición es que, en el caso en que m y n sean ambos números naturales, la definición de suma que acabamos de dar coincide con la que ya teníamos definida.

Es fácil convencerse de que esta definición algo técnica se corresponde con la idea geométrica que hemos discutido previamente: $m+n$ es el número al que llegamos si, partiendo de m, nos desplazamos n pasos hacia la derecha en el orden usual de los enteros si $n>0$, o bien $|n|$ pasos hacia la izquierda si $n<0$ (o no nos desplazamos si $n=0$).

Vamos probar las propiedades básicas de la suma de números enteros:[10]

Propiedad asociativa Dados tres números enteros, m, n, r, si $n\geq 0$ y $r\geq 0$, entonces $(m+n)+r$ es el número al que llegamos si nos desplazamos primero n lugares hacia la derecha y luego otros r lugares, pero claramente esto es lo mismo que desplazarse $n+r$ lugares hacia la derecha, es decir, $(m+n)+r=m+(n+r)$.

Si $n\leq 0$ y $r\leq 0$ razonamos análogamente, pero ahora considerando desplazamientos hacia la izquierda: desplazarse $|n|$ lugares hacia la izquierda y luego otros $|r|$ lugares es lo mismo que desplazarse $|n|+|r|$, lugares hacia la izquierda, pero éste es justo el efecto de sumar $-(|n|+|r|)=n+r$, es decir, de nuevo tenemos $(m+n)+r=m+(n+r)$.

Por último, si n y r tienen signos opuestos, entonces $(m+n)+r$ se obtiene de moverse $|n|$ pasos en un sentido y $|r|$ pasos en sentido opuesto, lo cual

[10] Los argumentos que siguen son las justificaciones de las propiedades que en la escuela se presentan sin justificación, de modo que el lector que esté familiarizado con su uso práctico sin conocer su justificación, puede optar por estudiarla ahora o seguir aceptando como algo conocido que se cumplen estos hechos básicos y saltarse los argumentos. Con las pruebas no aprenderá nada que no sepa ya.

equivale a moverse tantos pasos como la diferencia de los valores absolutos hacia la derecha si el mayor es el positivo o hacia la izquierda si el mayor es el negativo, pero esto es justo el efecto de sumar $n+r$, por la definición de suma, luego también en este caso $(m+n)+r = m+(n+r)$.

Propiedad conmutativa La conmutatividad de la suma es inmediata, pues en los tres casos en que hemos definido la suma el resultado es el mismo si se invierte el orden de los sumandos.

Compatibilidad con el orden Otro hecho inmediato es que si $m \leq n$, entonces $m + r \leq n + r$, pues si partimos de un número m situado antes que otro n y desde ambos realizamos el mismo movimiento, el punto al que llegamos desde el primero seguirá estando antes que el punto al que llegamos desde el segundo.

Producto de números enteros El producto de un número entero m por otro positivo $+n$ es el resultado de sumar n veces el número m, mientras que multiplicar por un número negativo $-n$ consiste en multiplicar por $+n$ y además cambiar el signo al resultado (y multiplicar por 0 da 0). Esto se plasma en la definición siguiente:

> *El producto $m \cdot n$ de dos números enteros m y n es el número entero cuyo valor absoluto es $|m \cdot n| = |m| \cdot |n|$ y cuyo signo (en el caso de que ambos factores sean no nulos) es positivo si ambos factores tienen el mismo signo y negativo si tienen signos opuestos.*

Nuevamente observamos que en el caso en que $m \geq 0$ y $n \geq 0$, el producto mn que acabamos de definir coincide con el producto de números naturales que ya teníamos definido.

En la práctica, si m es un número entero, escribiremos $-m$ en lugar de $-1 \cdot m$, que no es sino el número con el mismo valor absoluto y signo opuesto al de m.

Veamos las propiedades más relevantes del producto:

Propiedad asociativa El producto de números enteros es trivialmente asociativo, pues tanto $(mn)r$ como $m(nr)$ son números enteros de valor absoluto $|m||n||r|$ y cuyo signo es positivo si el número de factores negativos es par y negativo si el número de factores negativos es impar (o es cero si hay un factor nulo). Recordemos que dos números enteros son iguales si y sólo si tienen el mismo valor absoluto y el mismo signo.

Propiedad conmutativa La conmutatividad también es inmediata.

Propiedad distributiva Veamos que $m(n+r) = mn+mr$. En primer lugar, es fácil ver que se cumple para $m = -1$, es decir, que $-(n+r) = (-n)+(-r)$ o, equivalentemente, que si cambiamos el signo a los dos sumandos, la suma cambia de signo. Esto se sigue de la definición de suma.

A su vez, esto reduce la comprobación al caso en que $m > 0$, pues si $m = 0$ es trivial y si lo hemos probado en el caso $m > 0$ y tenemos $m < 0$, entonces $-m > 0$, luego tenemos que $-m(n+r) = (-mn)+(-mr)$, luego por el caso $m = -1$ concluimos que $m(n + r) = mn + mr$.

Suponemos, pues, que $m > 0$. Si $n \geq 0$ y $r \geq 0$, entonces se trata de la propiedad distributiva de los números naturales, que ya tenemos probada. Si $n \leq 0$ y $r \leq 0$, entonces, usando el caso $m = -1$ ya probado,

$$-m(n + r) = m((-n) + (-r)) = -mn + -nr = -(mn + mr),$$

luego $m(n + r) = mn + mr$.

Si n y r tienen signos opuestos, cambiándolos por $-n$ y $-r$ como en el caso anterior, podemos suponer que el positivo tiene mayor valor absoluto. Más aún, en virtud de la conmutatividad de la suma, no perdemos generalidad si suponemos que $n \geq 0$ y $r \leq 0$. Llamamos $x = n + r$, de manera que se cumple $x + (-r) = n$ es una igualdad entre números naturales, luego $mx+m(-r) = mn$, luego $mx-mr = mn$, luego $m(n+r) = mx = mn+mr$.

Anillos conmutativos y unitarios Ahora es inmediato que los números enteros constituyen lo que se conoce como un *dominio íntegro ordenado,* lo cual significa simplemente que satisfacen las propiedades recogidas en el resumen 1.5. Todas las propiedades que no hemos justificado se razonan trivialmente.

Dar nombres pintorescos como "anillo" o "dominio" a determinadas combinaciones de propiedades no es una frivolidad. Si el lector está familiarizado con la manipulación práctica de expresiones numéricas, pero no con estas estructuras teóricas, habrá dado un paso de gigante hacia la madurez necesaria para entender el comportamiento de los números en cuanto sea capaz de distinguir qué propiedades son válidas para operar con elementos de un anillo arbitrario, cuáles requieren estar en un dominio íntegro, cuáles dependen de la estructura de anillo ordenado, etc. Veamos algunos ejemplos de hechos elementales sobre números enteros que en realidad son válidos en anillos (conmutativos y unitarios) arbitrarios.[11] Aunque el lector conozca sobradamente estos hechos, debería aprender —si no lo sabe ya— que no sólo valen para números enteros, sino que pueden ser usados al trabajar en cualquier anillo, por diferente que sea del de los números enteros.

1. *Los elementos neutros que cumplen $a + 0 = a$ y $a \cdot 1 = a$ son únicos.*

   En efecto, si hubiera otro $0'$ con la misma propiedad, tendríamos que $0 = 0 + 0' = 0'$. Igualmente se razona con el 1.

2. *Para cada elemento a, el elemento opuesto $-a$ que cumple $a + (-a) = 0$ es único.*

[11] El concepto general de anillo no requiere que se cumplan ni la propiedad conmutativa para el producto ni la existencia de elemento neutro para el producto, pero nunca trabajaremos con anillos tan generales.

Resumen 1.5: Definición de anillo conmutativo y unitario

Un *anillo conmutativo y unitario* es un conjunto en el que hay definidas una suma y un producto que cumplan las propiedades siguientes:

Propiedad asociativa	$(a+b)+c = a+(b+c)$
Elemento neutro	Existe un elemento 0 tal que $a+0=a$.
Elemento opuesto	Para cada a existe un $-a$ tal que $a+(-a)=0$.
Propiedad conmutativa	$a+b=b+a$
Propiedad asociativa	$(ab)c = a(bc)$
Propiedad distributiva	$a(b+c) = ab+ac$
Elemento neutro	Existe un elemento 1 tal que $a\cdot 1 = a$.
Propiedad conmutativa	$ab = ba$

Un *dominio* es un anillo conmutativo y unitario tal que $1 \neq 0$.

Un *dominio íntegro* es un dominio que cumple además:

Integridad	Si $ab=0$, entonces $a=0$ o $b=0$.

Un anillo está *ordenado* si es un conjunto ordenado (Resumen 1.1) con una relación $\leq$ que además cumple las *relaciones de compatibilidad:*

con la suma	Si $b \leq c$ entonces $a+b \leq a+c$.
con el producto	Si $a \geq 0$ y $b \geq 0$, entonces $ab \geq 0$.

En efecto, si hubiera otro, digamos $-a'$, entonces

$$-a = -a+0 = -a+a+(-a') = 0+(-a') = -a'.$$

3. Sumandos iguales pueden simplificarse: de $m+r=n+r$ podemos pasar a $m=n$.

   Basta sumar $-r$ a ambos miembros para justificar el paso.

4. $a\cdot 0 = 0$ (multiplicar por 0 siempre da 0).

   En efecto:

   $$n\cdot 0 + n\cdot 0 = n\cdot(0+0) = n\cdot 0 = n\cdot 0 + 0,$$

   y cancelando el término común queda $n\cdot 0 = 0$.

Un *dominio* es un anillo conmutativo y unitario en el que $1 \neq 0$, pero con esta condición sólo estamos descartando un caso trivial, ya que si se cumple $1=0$, entonces todo elemento a del anillo cumple $a = a\cdot 1 = a\cdot 0 = 0$, luego resulta que el anillo no tiene más elementos que el 0.

Ejercicio: Razonar que $-(ab) = (-a)b = a(-b)$.

En un anillo arbitrario podemos definir la *resta* de dos elementos cualesquiera como

$$a - b = a + (-b).$$

Así, una resta en un anillo no es más que un caso particular de suma: la suma de un elemento con el opuesto de otro.

Más concretamente, $x = a - b$ es el único elemento que cumple $b + x = a$, por lo que en el caso concreto de los números enteros, tenemos que si m y n son dos números naturales con $m \leq n$, entonces la resta $n - m$ definida en el anillo de los números enteros coincide con la resta de números naturales que ya teníamos definida. Pero ahora también podemos restar $m - n = -(n - m)$, que es un número negativo (o 0).

En resumen, en el contexto de los números enteros podemos restar dos números naturales cualesquiera sin preocuparnos de averiguar cuál es mayor de los dos, y esta operación no es sino un caso particular de suma.

Dominios íntegros Los números enteros cumplen trivialmente que si $mn = 0$, entonces $m = 0$ o $n = 0$. Sin embargo, esta propiedad no es válida en dominios arbitrarios, luego a la hora de usarla (o de usar sus consecuencias) en un anillo distinto del de los enteros, tenemos que saber de antemano que se trata de un dominio íntegro. Una consecuencia relevante de la integridad es la posibilidad de simplificar factores no nulos:

Si $ab = ac$ y $a \neq 0$, entonces $b = c$.

En efecto, tenemos que $ab - ac = a(b - c) = 0$, luego por la integridad $b - c = 0$, es decir, $b = c$.

Potencias Si a es un elemento de un anillo y n es un número natural, podemos definir la potencia

$$a^n = \overbrace{a \cdots a}^{n \text{ veces}},$$

con el convenio de que $a^0 = 1$.

Este convenio se adopta para que la primera propiedad enunciada en el resumen 1.6 se cumpla también cuando $m = 0$ o $n = 0$. Así podemos aplicarla siempre sin preocuparnos de si r es o no 0.

Todas las propiedades enunciadas en el resumen 1.6 son triviales. Por ejemplo, en la tercera propiedad tenemos un producto con n factores iguales a ab, pero podemos verlo también como un producto con $2n$ factores, de los cuales n son iguales a a y otros n son iguales a b, luego tenemos $a^n b^n$. Igualmente se justifican las demás.

Multiplicación por enteros Es conocida la fórmula para el cuadrado de una suma:

$$(a+b)^2 = (a+b)(a+b) = a(a+b) + b(a+b) = a^2 + ab + ba + b^2 = a^2 + 2ab + b^2.$$

Resumen 1.6: Potencias en un anillo unitario

- $a^{m+n} = a^m \cdot a^n$
- $(a^m)^n = a^{mn}$
- $(ab)^n = a^n b^n$
- $1^n = 1$.
- $0^n = 0$, salvo si $n = 0$, en cuyo caso $0^0 = 1$.

Este desarrollo es válido en un anillo conmutativo arbitrario, pero en él hemos introducido una notación que no tenemos justificada en general, y es $2ab$. Esto tiene sentido si consideramos que a y b son números enteros, pero un anillo arbitrario no tiene por qué contener el número entero 2, por lo que en principio no podemos calcular $2ab$. Ahora bien, es fácil definir el producto de un número entero n por un elemento a de un anillo arbitrario:

$$na = \begin{cases} \overbrace{a + \cdots + a}^{n \text{ veces}} & \text{si } n > 0, \\ 0 & \text{si } n = 0, \\ \underbrace{-a - \cdots - a}_{-n \text{ veces}} & \text{si } n < 0. \end{cases}$$

Si a es también un número entero, es claro que na es el producto usual de números enteros, pero ahora podemos decir que $a + a = 2a$ en cualquier anillo. El resumen 1.7 contiene las propiedades básicas de este producto por enteros.

Dominios ordenados Veamos ahora algunas propiedades que son válidas en un dominio ordenado arbitrario:

1. Si $a < b$, entonces $a + c < b + c$.

   Por definición de anillo ordenado tenemos $a + c \leq b + c$, pero si fuera $a + c = b + c$ podríamos simplificar y concluir que $a = b$.

Resumen 1.7: Multiplicación por enteros

- $n(a + b) = na + nb$,
- $(m + n)a = ma + na$,
- $m(na) = (mn)a$,
- $m(ab) = (ma)b$.
- $1 \cdot a = a$.

2. $a \leq 0$ si y sólo si $-a \geq 0$.

   Si $a \leq 0$, la compatibilidad con la suma nos da que $0 = a - a \leq 0 - a = -a$. El recíproco se prueba análogamente.

3. $a^2 \geq 0$.

   En efecto, si $a \geq 0$ esto es un caso particular de la relación de compatibilidad con el producto, mientras que si $a \leq 0$, entonces $-a \geq 0$, luego la compatibilidad con el producto nos da que $0 \leq (-a)^2 = a^2$.

4. $1 > 0$.

   Porque $1 = 1 \cdot 1$.

5. $-1 < 0$

   Por 1) y 3).

6. Si $a \geq 0$ y $b \leq c$, entonces $ab \leq ac$.

   La compatibilidad con la suma nos da que $c - b \geq 0$, y la compatibilidad con el producto que $a(c - b) \geq 0$, luego $ac - ab \geq 0$ y, de nuevo por la compatibilidad con la suma, concluimos que $ab \leq cd$.

7. Si $a \leq 0$ y $b \leq c$, entonces $ab \geq ac$.

   Si $a \leq 0$, entonces $-a \geq 0$, luego $-ab \leq -ac$, luego $ac \leq ab$.

En un dominio íntegro ordenado se cumple además la relación de compatibilidad del producto con el orden estricto:

Si $a > 0$ y $b > 0$, entonces $ab > 0$.

En efecto, en principio tiene que cumplirse $ab \geq 0$ por definición de anillo ordenado y $ab \neq 0$ por definición de dominio íntegro. De aquí se siguen a su vez las propiedades análogas a 6) y 7) para desigualdades estrictas.

Hay una propiedad notable de los dominios ordenados:

Todo dominio ordenado contiene a los números enteros.

En efecto, en un dominio ordenado se cumple que $-1 < 0 < 1$ y, a partir de aquí, sumando 1 y -1 obtenemos que

$$\cdots < -1 - 1 - 1 < -1 - 1 < -1 < 0 < 1 < 1 + 1 < 1 + 1 + 1 < \cdots$$

por lo que si, para cada entero n, llamamos $\bar{n} = n \cdot 1$, tenemos que

$$\cdots < \overline{-3} < \overline{-2} < \overline{-1} < \bar{0} < \bar{1} < \bar{2} < \bar{3} \cdots$$

son elementos del dominio ordenado considerado, y las propiedades del resumen 1.7 nos dan que

$$\bar{m} + \bar{n} = \overline{m + n}, \qquad \bar{m} \cdot \bar{n} = \overline{mn}.$$

Esto quiere decir que se cumple, por ejemplo, $\bar{3} + \overline{-5} = \overline{-2}$, $\bar{3} \cdot \overline{-5} = \overline{-15}$ y, en general, que podemos identificar los elementos $\bar{n}$ del dominio con los números enteros, pues están ordenados igual, se suman igual y se multiplican igual.

En la práctica podemos omitir las barras y entender que, en todo dominio ordenado, los números que se obtienen sumando 1 y -1 muchas veces no son sino los números enteros.

El *valor absoluto* puede definirse en cualquier anillo ordenado como

$$|a| = \begin{cases} a & \text{si } a \geq 0, \\ -a & \text{si } a \leq 0. \end{cases}$$

Es claro que en el caso de los números enteros esta definición nos da el valor absoluto que ya tenemos definido. Veamos algunas propiedades:

1. $|a| \leq b$ si y sólo si $-b \leq a \leq b$.

   En efecto, si $|a| \leq b$ y $a \geq 0$, entonces $a = |a| \leq b$, luego $-b \leq -a \leq 0 \leq a$. Si es $a \leq 0$, entonces $-a = |a| \leq b$, luego $-b \leq a \leq 0 \leq |a| \leq b$.

   Recíprocamente, si $-b \leq a \leq b$, entonces, también $-a \leq b$, luego $|a| \leq b$, pues o bien $|a| = a$ o bien $|a| = -a$.

2. $|a + b| \leq |a| + |b|$.

   Basta aplicar la propiedad anterior a $|a| \leq |a|$, con lo que obtenemos que $-|a| \leq a \leq |a|$ e igualmente $-|b| \leq b \leq |b|$, luego

   $$-|a| - |b| \leq a + b \leq |a| + |b|,$$

   luego $|a + b| \leq |a| + |b|$.

3. $||a| - |b|| \leq |a - b|$.

   En efecto, $|a| = |a - b + b| \leq |a - b| + |b|$, luego $|a| - |b| \leq |a - b|$. Intercambiando los papeles de a y b y teniendo en cuenta que

   $$|b - a| = |-(a - b)| = |a - b|,$$

   obtenemos que $|b| - |a| \leq |a - b|$ o, equivalentemente,

   $$-|a - b| \leq |a| - |b| \leq |a - b|,$$

   de donde $||a| - |b|| \leq |a - b|$.

El Resumen 1.8 contiene las propiedades fundamentales del valor absoluto. Las que no hemos demostrado son inmediatas.

Resumen 1.8: El valor absoluto en un anillo ordenado

- $|a| = 0$ si y sólo si $a = 0$.
- $|ab| = |a||b|$
- $|a + b| \leq |a| + |b|$

Como consecuencia:

$$||a| - |b|| \leq |a - b|.$$

1.3 Polinomios

En la sección anterior hemos razonado que algunas propiedades básicas de los números enteros son válidas en general en dominios arbitrarios, o en dominios íntegros arbitrarios, etc. Es razonable que el lector no familiarizado con las estructuras algebraicas abstractas se pregunte con recelo si realmente merece la pena enunciar que multiplicar por 0 da 0 en un anillo arbitrario, en lugar de ahorrarnos la jerga algebraica y considerar únicamente el caso numérico, que es el que nos interesa.

Para aliviar esos posibles recelos interrumpimos la revisión del sistema numérico dejando para la sección siguiente la presentación de los números racionales y estudiamos aquí un ejemplo de una familia de anillos cuyo interés para el estudio de los números es fácil de entrever y que ilustra la utilidad de enunciar los hechos básicos en términos de anillos arbitrarios en la medida de lo posible. Se trata de los anillos de polinomios, que son expresiones de la forma

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0,$$

donde x es una *cantidad indeterminada* y los coeficientes a_i están en un dominio prefijado.

Esto plantea el problema "filosófico" de qué queremos decir exactamente con que x es "una cantidad indeterminada", pero la definición siguiente elimina el problema de raíz:

> Un *polinomio* con coeficientes en un dominio A es una sucesión infinita
>
> $$P = \quad a_0, \quad a_1, \quad a_2, \quad a_3, \quad \ldots$$
>
> de elementos de A cuyos términos son todos nulos a partir de uno dado. Se dice que a_i es el *coeficiente de grado i* del polinomio. El polinomio con todos los coeficientes nulos se llama *polinomio nulo,* y el *grado* de un polinomio no nulo P se define como el mayor índice n tal que $a_n \neq 0$. Dicho coeficiente a_n se llama *coeficiente director* de P. Consideraremos también que el polinomio nulo tiene grado 0.

Tenemos así una definición de polinomio sin "cantidades indeterminadas", pero a continuación las introducimos "de contrabando" con el convenio siguiente:

Si

$$P = \quad a_0, \quad a_1, \quad a_2, \quad \ldots \quad a_{n-1}, \quad a_n, \quad 0, \quad 0, \quad \ldots$$

es un polinomio con coeficientes en un dominio A (donde no exigimos que $a_n \neq 0$, sino únicamente que los coeficientes posteriores sean nulos), lo representaremos en la forma

$$\sum_{i=0}^{n} a_i x^i = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x^1 + a_0 x^0.$$

Más aún, en la práctica, al escribir explícitamente un polinomio, escribiremos $a_1 x$ en lugar de $a_1 x^1$, a_0 en lugar de $a_0 x^0$ y omitiremos los $a_i = 1$, así como los términos $a_i x^i$ cuando $a_i = 0$. En particular, la indeterminada x no es sino el polinomio

$$x = \quad 0, \quad 1, \quad 0, \quad 0, \quad \ldots$$

Por ejemplo,

$$8x^5 + x^3 - 7x^2 + 5x$$

no es sino una forma de representar la sucesión de números enteros

$$0, \quad 5, \quad -7, \quad 1, \quad 0, \quad 8, \quad 0, \quad 0, \quad \ldots$$

En principio estamos obligados a respetar el orden de los índices, de modo que esto:

$$-7x^2 + 5x + 8x^5 + x^3$$

no es un polinomio, pero esta restricción desaparece en cuanto consideramos la definición siguiente de suma de polinomios:

$$\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i = \sum_{i=0}^{n} (a_i + b_i) x^i,$$

es decir, que dos polinomios[12] se suman sumando los coeficientes de cada grado. Por ejemplo:

$$\begin{array}{rccccccc} & 8x^5 & & +x^3 & -7x^2 & +5x & \\ + & & 2x^4 & -x^3 & +x^2 & & +3 \\ \hline & 8x^5 & +2x^4 & & -6x^2 & +5x & +3 \end{array}$$

Es obvio que la suma de polinomios es asociativa, conmutativa, tiene por elemento neutro al polinomio nulo y todo elemento tiene un simétrico (el que

[12] Notemos que no perdemos generalidad al pedir que i llegue hasta n en ambos sumandos, pues n se puede tomar arbitrariamente grande incluyendo coeficientes $a_i = 0$ en la representación.

resulta de sustituir cada coeficiente por su simétrico). Por ejemplo, la propiedad asociativa se demuestra así:

$$\left(\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i\right) + \sum_{i=0}^{n} c_i x^i = \sum_{i=0}^{n} (a_i + b_i)x^i + \sum_{i=0}^{n} c_i x^i =$$

$$\sum_{i=0}^{n} (a_i + b_i + c_i)x^i = \sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} (b_i + c_i)x^i = \sum_{i=0}^{n} a_i + \left(\sum_{i=0}^{n} b_i x^i + \sum_{i=0}^{n} c_i x^i\right).$$

Así pues, al cumplirse la propiedad asociativa, podemos escribir sumas con cualquier número de sumandos sin necesidad de poner paréntesis, y ahora podemos afirmar que[13]

$$-7x^2 + 5x + 8x^5 + x^3 = 8x^5 + x^3 - 7x^2 + 5x$$

Los polinomios de la forma $a_i x^i$ se llaman *monomios,* y ahora es inmediato que todo polinomio no nulo se expresa de forma única (salvo el orden) como suma de monomios no nulos. Ahora vamos a definir un producto de polinomios que nos permita interpretar un monomio ax^i como el producto del polinomio

$$a = \quad a, \quad 0, \quad 0, \quad \ldots$$

por i veces el polinomio x.

La definición del producto es la siguiente:

$$\left(\sum_{i=0}^{m} a_i x^i\right)\left(\sum_{i=0}^{n} b_i x^i\right) = \sum_{i=0}^{m+n} \Big(\sum_{u+v=i} a_u b_v\Big) x^i.$$

Equivalentemente, el coeficiente de grado i del producto es la suma de todos los productos de coeficientes, uno de cada factor, cuyos grados suman i. De nuevo es fácil probar que este producto es asociativo y conmutativo. Por ejemplo, la propiedad asociativa se demuestra así:

$$\left(\left(\sum_{i=0}^{m} a_i x^i\right)\left(\sum_{i=0}^{n} b_i x^i\right)\right)\left(\sum_{i=0}^{p} c_i x^i\right) = \left(\sum_{i=0}^{m+n} \Big(\sum_{u+v=i} a_u b_v\Big) x^i\right)\left(\sum_{i=0}^{p} c_i x^i\right)$$

$$= \sum_{i=0}^{m+n+p} \left(\sum_{t+w=i} \sum_{u+v=t} a_u b_v\Big) c_w\right) x^i = \sum_{i=0}^{m+n+p} \Big(\sum_{u+v+w=i} a_u b_v c_w\Big) x^i,$$

y si partimos de los factores asociados al revés, llegamos a la misma expresión final.

Ahora observamos que el producto del polinomio x^n por el polinomio x^m es el polinomio x^{n+m}, pues $x^n = \sum_{i=0}^{n} a_i x^i$ con todos los $a_i = 0$ salvo $a_n = 1$, e igualmente $x^m = \sum_{i=0}^{m} b_i x^i$, con todos los $b_i = 0$ salvo $b_m = 1$, luego el único producto $a_i b_j$ no nulo es $a_n b_m = 1$, luego el producto es x^{n+m}.

[13]Como curiosidad técnica, insistimos en que esta igualdad no requiere la propiedad conmutativa, sino que de hecho es la propia definición de suma, pues, por definición se cumple que $(-7x^2 + 5x) + (8x^5 + x^3) = (0 + 8)x^5 + (0 + 1)x^3 + (-7 + 0)x^2 + (5 + 0)x$.

Por lo tanto, el producto del polinomio x por sí mismo n veces es precisamente x^n, luego a partir de ahora ya podemos considerar que el polinomio que hemos llamado x^m no es sino x elevado a m en el sentido algebraico usual.

Más aún, al multiplicar $a = \sum_{i=0}^{0} a_i x^i$, con $a_0 = a$, por $x^n = \sum_{i=0}^{n} b_i x^i$, con $b_i = 0$ salvo $b_n = 1$, el único producto $a_i b_j$ no nulo es $a_0 b_n = a$, luego el resultado es ax^n.

Con esto ya tenemos justificado que en toda expresión de un polinomio en la forma

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0,$$

las sumas son sumas, los productos son productos y las potencias son potencias. Las propiedades asociativa y conmutativa nos permiten olvidarnos de la fórmula con la que hemos definido el producto de polinomios y reducir el cálculo de cualquier producto a sumas de productos de monomios, y el producto de dos monomios se reduce a multiplicar sus coeficientes (con el producto del dominio A) y multiplicar las potencias de x sumando los exponentes. En la práctica, esto significa que podemos multiplicar polinomios aplicando el mismo algoritmo que empleamos para multiplicar números enteros a partir de sus expresiones decimales. Por ejemplo:

$$\begin{array}{rrrrrr}
 & & 3x^3 & +5x^2 & -2x & +7 \\
 & \times & & -2x^2 & & +2 \\
\hline
 & & 6x^3 & +10x^2 & -4x & +14 \\
-6x^5 & -10x^4 & +4x^3 & -14x^2 & & \\
\hline
-6x^5 & -10x^4 & +10x^3 & -4x^2 & -4x & +14
\end{array}$$

En particular sucede que los polinomios de grado 0 se pueden identificar con los elementos de A, pues se suman y se multiplican como los elementos de A. También es inmediato que el polinomio 1 es el elemento neutro para el producto de polinomios.

Anillos de polinomios Como conclusión de la discusión previa podemos afirmar lo siguiente:

> *Si A es un dominio, el conjunto $A[x]$ de todos los polinomios con coeficientes en A es también un dominio. Todo polinomio no nulo con coeficientes en A se expresa en la forma*
>
> $$P = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$
>
> *y la expresión es única si exigimos que $a_n \neq 0$ sea el coeficiente director del polinomio.*

A partir de aquí, es decir, una vez tenemos garantizado que los polinomios forman un dominio que contiene al dominio A de partida, podemos "olvidarnos" de toda la construcción que hemos hecho, es decir, podemos olvidarnos de que

hemos definido los polinomios como sucesiones de coeficientes, a condición de que nos comprometamos a no hacernos preguntas "filosóficas" como ¿cuánto vale la indeterminada x? Ante esta pregunta, o bien somos prácticos y la consideramos una impertinencia, o bien nos ponemos formalistas y recordamos que x no es más que la sucesión $0, 1, 0, 0, \ldots$, luego no es nada que pueda tomar ningún valor.

En particular, debemos distinguir una indeterminada de un anillo de polinomios de una "incógnita" cuyo valor queremos calcular. Por ejemplo, si nos planteamos si $x^2-4=0$ y entendemos el miembro izquierdo como un polinomio, la respuesta no es que la igualdad se da cuando $x=\pm 2$, sino que la igualdad no se da nunca, porque el miembro izquierdo es un polinomio de grado 2 que no es el polinomio nulo, se mire como se mire. Lo sería si fuera $x=\pm 2$, ciertamente, pero es que $x \neq \pm 2$, porque x es un polinomio de grado 1 y ± 2 son polinomios de grado 0.

Sustitución Otra cosa distinta es que, dado un polinomio, podemos calcular el elemento que resulta al sustituir su indeterminada por un elemento del dominio A (o de cualquier otro dominio que lo contenga). Más precisamente, es habitual usar la notación

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

para representar un polinomio arbitrario, y entonces, para cada elemento c de A (o de cualquier dominio B que contenga a A), podemos definir la sustitución

$$f(c) = a_n c^n + a_{n-1}c^{n-1} + \cdots + a_2 c^2 + a_1 c + a_0,$$

que es un elemento de B, y no es difícil convencerse de que

$$(f+g)(c) = f(c) + g(c), \qquad (fg)(c) = f(c)g(c).$$

La razón de fondo es que cuando sumamos y multiplicamos polinomios no hacemos nada que no podríamos hacer si x fuera cualquier elemento de cualquier dominio, por lo que todas las operaciones siguen siendo válidas si sustituimos x por cualquier elemento de cualquier dominio que contenga a A.

El grado de un polinomio Si representamos por $\operatorname{grad} f(x)$ el grado del polinomio $f(x)$, es claro que

$$\operatorname{grad}(f(x)+g(x)) \leq \operatorname{máx}\{\operatorname{grad} f(x), \operatorname{grad} g(x)\}.$$

No tiene por qué darse la igualdad, ya que los monomios de mayor grado se pueden cancelar. Por ejemplo:

$$(4x^3 - 3x + 5) + (-4x^3 + x + 2) = -2x + 7$$

es una suma de dos polinomios de grado 3 que resulta tener grado 1. No obstante, si los dos sumandos tienen grado distinto, entonces el término de mayor grado del de grado mayor no puede cancelarse y tenemos la igualdad.

Para el producto sucede algo similar: $\operatorname{grad}(f(x)g(x)) \leq \operatorname{grad} f(x)+\operatorname{grad} g(x)$, pero la igualdad falla si uno de los factores es nulo, pues entonces el producto también es nulo. Si los factores son no nulos y los monomios de mayor grado son, respectivamente, ax^m y bx^n, con $a \neq 0 \neq b$, al calcular el producto aparecerá abx^{m+n} como único monomio de grado $m+n$, que no podrá cancelarse con ningún otro, pero este monomio podría ser nulo si $ab = 0$, con lo que el grado del producto puede ser inferior a $m+n$.

Naturalmente, esto no sucede si el dominio de partida es íntegro. En tal caso, el monomio abx^{m+n} es no nulo y concluimos que el producto de dos polinomios no nulos cumple

$$\operatorname{grad}(f(x)g(x)) = \operatorname{grad} f(x) + \operatorname{grad} g(x).$$

En particular:

> *Si A es un dominio íntegro, entonces el anillo de polinomios $A[x]$ también es un dominio íntegro.*

En efecto, acabamos de probar que el producto de dos polinomios no nulos no puede ser nulo, pues contiene el monomio no nulo abx^{m+n}.

Ejemplo Consideremos las potencias cuartas de los primeros números naturales:

$$1,\quad 16,\quad 81,\quad 256,\quad 625,\quad 1\,296,\quad 2\,401,\quad 4\,096,\quad 6\,561,\quad 10\,000.$$

Ahora calculemos las diferencias entre cada número obtenido y su anterior:

$$15,\quad 65,\quad 175,\quad 369,\quad 671,\quad 1\,105,\quad 1\,695,\quad 2\,465,\quad 3\,439.$$

Otra vez:

$$50,\quad 110,\quad 194,\quad 302,\quad 434,\quad 590,\quad 770,\quad 974.$$

Y otra vez:

$$60,\quad 84,\quad 108,\quad 132,\quad 156,\quad 180,\quad 204.$$

Y a la cuarta vez obtenemos

$$24,\quad 24,\quad 24,\quad 24,\quad 24,\quad 24.$$

Si el lector parte de otro exponente distinto de cuatro llegará a un resultado similar. ¿Podría el lector razonar por qué sucede esto y cuál es el número al que se llega finalmente?

En lugar de partir de potencias x^n, vamos a considerar un polinomio arbitrario $p(x)$ de grado n no nulo. Entonces $p(x+1) - p(x)$ es un polinomio de grado $n-1$ cuyo coeficiente director es n veces el coeficiente director de $p(x)$.

En efecto, sea $p(x) = \sum_{i=0}^{n} a_i x^i$. Así

$$p(x+1) - p(x) = \sum_{i=0}^{n} a_i (x+1)^i - \sum_{i=0}^{n} a_i x^i.$$

Cada polinomio $(x+1)^i$ tiene grado i, luego el único monomio de grado n que aparece en $p(x+1)$ es el de $a_n(x+1)^n$, o sea, $a_n x^n$, que se anula con el monomio correspondiente de $p(x)$, luego el grado de $p(x+1)-p(x)$ es a lo sumo $n-1$. Ahora calculemos el monomio de grado $n-1$.

En $\sum_{i=0}^{n} a_i(x+1)^i$ tenemos dos sumandos con grado $\geq n-1$, a saber, $a_n(x+1)^n$ y $a_{n-1}(x+1)^{n-1}$. El monomio de grado $n-1$ en cada uno de ellos es $na_n x^{n-1}$ y $a_{n-1}x^{n-1}$, respectivamente, pero el último se cancela con el correspondiente monomio de $p(x)$. Por tanto el monomio de grado $n-1$ en $p(x+1)-p(x)$ es exactamente $na_n x^{n-1}$, con lo que ciertamente se trata de un polinomio de grado $n-1$ con coeficiente director na_n.

En consecuencia, si partimos del polinomio x^n, las diferencias sucesivas son los valores que toma el polinomio $(x+1)^n - x^n$, que es un polinomio de grado $n-1$ con coeficiente director n, las siguientes diferencias vienen dadas por un polinomio de grado $n-2$ y coeficiente director $n(n-1)$, luego las diferencias n–simas vienen dadas por un polinomio de grado 0, o sea, constante y con coeficiente director igual al *factorial* de n, definido como:

$$n! = n(n-1)(n-2)\cdots 2\cdot 1,$$

luego todas las diferencias n-simas son iguales a $n!$ ∎

División euclídea Es posible dividir euclídeamente los polinomios de forma análoga a como dividimos números naturales. En realidad hace falta un pequeño requisito para que una división euclídea pueda realizarse.

En general, se dice que un elemento a de un dominio A es una *unidad* si existe un a^{-1} en A tal que $a\cdot a^{-1}=1$.

Por ejemplo, es fácil ver que en el dominio de los números enteros las únicas unidades son ± 1.

> *Si $D(x)$ y $d(x)$ son polinomios con coeficientes en un dominio íntegro A, $d(x)$ es no nulo y su coeficiente director es una unidad, entonces existen unos únicos polinomios $c(x)$ y $r(x)$ con coeficientes en A tales que $D(x) = d(x)c(x) + r(x)$ con $r(x) = 0$ o bien* $\operatorname{grad} r(x) < \operatorname{grad} d(x)$.

Observemos que la condición sobre el resto se podría reducir simplemente[14] a $\operatorname{grad} r(x) < \operatorname{grad} d(x)$ salvo en el caso en que el divisor $d(x)$ tiene grado 0, pero entonces $d(x) = a$, donde a es, por hipótesis una unidad de A, y entonces basta tomar $c(x) = a^{-1}D(x)$ y $r(x) = 0$.

[14]Hay un truco para no tener que estar tratando aparte al polinomio nulo cada vez que hablamos de grados, y consiste en convenir que $\operatorname{grad} 0 = -\infty$, entendiendo que $-\infty$ es un valor menor que todo número entero y que $-\infty + n = -\infty$ para todo número entero. Con estos convenios, los polinomios sobre dominios íntegros cumplen que

$$\operatorname{grad}(p(x)+q(x)) \leq \operatorname{máx}\{\operatorname{grad} p(x), \operatorname{grad} q(x)\}, \quad \operatorname{grad}(p(x)q(x)) = \operatorname{grad} p(x) + \operatorname{grad} q(x)$$

y, en la división euclídea, $\operatorname{grad} r(x) < \operatorname{grad} d(x)$, sin necesidad de tratar aparte al polinomio nulo.

Para justificar que es posible dividir polinomios en estas condiciones vamos a considerar un ejemplo en concreto, pero el lector podrá apreciar que todo se puede hacer en general con polinomios arbitrarios con coeficientes en dominios íntegros arbitrarios. Más aún, la integridad del dominio sólo es necesaria para justificar la unicidad del cociente y del resto.

Vamos a mostrar con un ejemplo concreto cómo podemos calcular el cociente y el resto de una división, pero será evidente que el procedimiento se puede aplicar para cualquier par de polinomios (con tal de que el coeficiente director del divisor sea un polinomio). Tomamos

$$D(x) = 3x^5 - 7x^4 - 9x^3 + 14x^2 - 13x - 3, \qquad d(x) = x^2 - 3x.$$

Observemos que el coeficiente director del divisor es 1, luego es una unidad. Notamos también que $\operatorname{grad} D(x) \geq \operatorname{grad} d(x)$. Si no fuera así, bastaría tomar $c(x) = 0$ y $r(x) = D(x)$ y ya estaría la división hecha. La división se realiza así:

$$\begin{array}{rrrrrrr|l}
3x^5 & -7x^4 & -9x^3 & +14x^2 & -13x & -3 & x^2 \quad -3x \\
\cline{8-8}
3x^5 & -9x^4 & & & & & 3x^3 +2x^2 -3x +5 \\
\cline{1-2}
 & 2x^4 & -9x^3 & & & & \\
 & 2x^4 & -6x^3 & & & & \\
\cline{2-3}
 & & -3x^3 & +14x^2 & & & \\
 & & -3x^3 & +\ 9x^2 & & & \\
\cline{3-4}
 & & & 5x^2 & -13x & & \\
 & & & 5x^2 & -15x & & \\
\cline{4-5}
 & & & & 2x & -3 &
\end{array}$$

En general, si el monomio de mayor grado del dividendo es cx^n y el del divisor es ax^m, donde a es una unidad y $m \leq n$, tomamos como monomio de mayor grado del cociente $c_1(x) = a^{-1}cx^{n-m}$, en nuestro caso $3x^3$. Así, si multiplicamos $a^{-1}cx^{n-m}d(x)$ obtenemos un polinomio cuyo monomio de mayor grado es cx^n, luego al restárselo a $D(x)$ resulta un polinomio $D_1(x)$ tal que $\operatorname{grad} D_1(x) < \operatorname{grad} D(x)$ que cumple $D(x) = d(x)c_1(x) + D_1(x)$.

En nuestro caso $D_1(x) = 2x^2 - 9x^3 + 14x^2 - 13x - 3$, pero, por simplicidad, sólo vamos bajando cada monomio cuando lo vamos necesitando. Si fuera $\operatorname{grad} D_1(x) < \operatorname{grad} d(x)$ ya estaría la división terminada, pero como no es así, repetimos el proceso dividiendo $D_1(x)$ entre $d(x)$, para obtener un nuevo cociente $c_2(x)$ y un nuevo resto $D_2(x)$ de grado menor tal que

$$D_1(x) = d(x)c_2(x) + D_2(x).$$

Así,

$$D(x) = d(x)c_1(x) + d(x)c_2(x) + D_2(x) = d(x)(c_1(x) + c_2(x)) + D_2(x).$$

En nuestro ejemplo, $D_2(x) = -3x^3 + 14x^2 - 13x - 3$. Si este polinomio ya tuviera grado menor que el de $d(x)$, la división estaría terminada, y si no es así, repetimos el proceso para obtener polinomios tales que

$$D(x) = d(c)(c_1(x) + c_2(x) + c_3(x)) + D_3(x).$$

Como cada polinomio $D_i(x)$ tiene grado menor que el anterior, tras un número finito de pasos tenemos que llegar a un resto $D_n(x)$ de grado menor que $d(x)$, y en este punto hemos terminado la división. En nuestro ejemplo llegamos a que

$$c(x) = 3x^3 + 2x^2 - 3x + 5, \qquad r(x) = 2x - 3.$$

Queda claro que este proceso puede llevarse a cabo con polinomios cualesquiera. La unicidad es fácil de probar: si tuviéramos

$$d(x)c_1(x) + r_1(x) = d(x)c_2(x) + r_2(x)$$

con restos nulos o de grado menor que el del divisor, entonces

$$d(x)(c_1(x) - c_2(x)) = r_2(x) - r_1(x),$$

con $\operatorname{grad}(r_2(x) - r_1(x)) < \operatorname{grad} d(x)$, mientras que el grado del miembro derecho es $\geq \operatorname{grad} d(x)$ a menos que $c_1(x) - c_2(x) = 0$, luego tiene que ser $c_1(x) = c_2(x)$ y esto implica a su vez que $r_1(x) = r_2(x)$.

Polinomios con varias indeterminadas Para terminar esta sección observamos que hasta ahora hemos considerado únicamente polinomios con una indeterminada, pero nada impide considerar igualmente polinomios con varias indeterminadas, como

$$5x^5y^2 + 3x^2y^3 + 2y^2 - 2xy + 7x - 5y + 7.$$

Una construcción directa de un dominio $A[x, y]$ resulta un poco más farragosa de explicitar, pero podemos evitarlo definiendo $A[x, y] = A[x][y]$. En efecto, dado un dominio A, tenemos que $A[x]$ es otro dominio, luego tenemos definido el anillo de los polinomios con coeficientes en $A[x]$ y con una indeterminada y, y a este anillo lo podemos llamar $A[x, y]$. Por ejemplo, si A es el anillo de los números enteros, un elemento de $A[x][y]$ es de la forma

$$(3x^2)y^3 + (5x^5 + 2)y^2 - (2x + 5)y + (7x + 7).$$

Se trata de un polinomio de grado 3 con coeficientes en $A[x]$, pero si aplicamos la propiedad distributiva observamos que es el mismo que habíamos escrito un poco más arriba. En general, es claro que todo polinomio de $A[x, y]$ se puede expresar de forma única como suma de monomios de la forma ax^iy^j, siempre y cuando no repitamos pares de exponentes (i, j).

Sabemos que si A es un dominio, también lo es $A[x, y]$, y del mismo modo podemos considerar anillos de polinomios $A[x_1, \ldots, x_n]$ con cualquier número de indeterminadas.[15]

[15] El lector con pensamientos filosóficos morbosos tal vez se pregunte con qué derecho escribimos $A[x][y]$ en lugar de $A[x][x]$ o, equivalentemente, por qué podemos afirmar que $x \neq y$, dado que una indeterminada no es más que la sucesión $0, 1, 0, 0, \ldots$, y ahí no hay nada que justifique la distinción entre x e y. La respuesta es que $x = 0, 1, 0, \ldots$, donde 1 es el elemento neutro del producto de A, mientras que $y = 0, 1, 0, \ldots$, donde 1 es el elemento neutro del producto de $A[x]$, es decir, la sucesión $1, 0, 0, \ldots$, luego $x \neq y$.

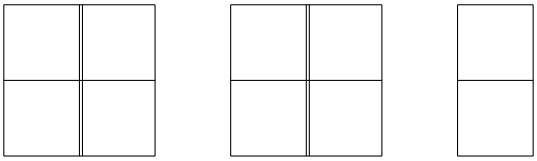
1.4 Los números racionales

Si los números naturales son la herramienta matemática para "contar", los números racionales son la herramienta matemática básica para "medir". Por ejemplo, si queremos medir un lapso de tiempo usando como unidad la "hora", no podemos esperar que el tiempo que queremos medir se ajuste a nuestra unidad y resulte ser exactamente de una hora, o de dos horas, o de tres horas, etc. Por el contrario, es perfectamente posible que sea de media hora, o de un cuarto de hora, o de una hora y 19 minutos (es decir, $60 + 19 = 79$ sesentavas partes de una hora). Este tipo de situaciones nos llevan a considerar, no únicamente múltiplos de una unidad dada, sino también cantidades fraccionarias.

Fracciones Desde un punto de vista puramente formal, la idea subyacente al concepto de número racional es que, del mismo modo que los números enteros nos permiten restar números naturales sin preocuparnos de cuál es mayor, los números racionales nos permiten dividir números naturales cualesquiera sin más precaución que la de no dividir entre 0, pero sin necesidad de que el divisor divida realmente al dividendo.

En particular, esto requiere que una unidad pueda dividirse en tantas partes como queramos, y esto nos lleva al concepto de fracción. Por ejemplo, $1/4$ representa el resultado de dividir una unidad en cuatro partes iguales y tomar una de ellas, mientras que $10/4$ representa a 10 de tales cuartas partes. Esta expresión recibe el nombre de *fracción* de numerador 10 y denominador 4. El numerador de una fracción puede tomar cualquier valor, pero no tiene sentido que el denominador sea 0 (no podemos dividir una unidad en 0 partes).

Una dificultad técnica que surge al tratar con fracciones es que dos fracciones con distintos numeradores y denominadores pueden representar la misma cantidad. Por ejemplo, tiene que ser $10/4 = 5/2$:



En general, es fácil convencerse de que tiene que ser

$$\frac{a}{b} = \frac{ac}{bc}$$

por lo que, a la hora de comparar dos fracciones podemos reducirlas a un denominador común:

$$\frac{a}{b} = \frac{ad}{bc} \qquad \frac{bc}{bd} = \frac{c}{d}$$

y observar que para que dos fracciones con el mismo denominador sean iguales es necesario y suficiente que sus numeradores sean iguales ($10/4$ no puede ser igual a ninguna cantidad de cuartos que no sea precisamente 10 cuartos). Concluimos

que la condición necesaria y suficiente para que dos fracciones representen la misma cantidad es

$$\frac{a}{b} = \frac{c}{d} \qquad \text{si y sólo si} \qquad ad = bc.$$

Éste es el punto de partida de la construcción de los números racionales. Ahora bien, conviene observar que toda la construcción puede hacerse sin modificación alguna considerando, no fracciones de números enteros, sino fracciones de elementos de un dominio íntegro arbitrario. Así, por ejemplo, podremos hablar igualmente de lo que se conoce como fracciones algebraicas, es decir, fracciones de polinomios, como

$$\frac{3x5 + 4x^3 + 1}{x^3 - 1}$$

Llamaremos *fracciones* en un dominio íntegro dado a las expresiones de la forma

$$\frac{a}{b},$$

donde a y b son elementos del dominio íntegro y $b \neq 0$, con el convenio de que dos fracciones son iguales si satisfacen la relación:

$$\frac{a}{b} = \frac{c}{d} \quad \text{si y sólo si} \quad ad = bc.$$

Dada una fracción a/b, se dice que a es su *numerador* y b su *denominador.*

Las fracciones de números enteros se llaman *números racionales.* Las fracciones de anillos de polinomios $A[x_1, \ldots, x_n]$ con coeficientes en un dominio íntegro A se llaman *fracciones algebraicas* con coeficientes en A con n indeterminadas. El conjunto de todas ellas se representa por $A(x_1, \ldots, x_n)$.

Observemos que

$$\frac{a}{1} = \frac{c}{1} \quad \text{si y sólo si} \quad a = c,$$

por lo que podemos identificar cada elemento a del dominio íntegro de partida con la fracción $a/1$. Definimos como sigue la suma y el producto de fracciones:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.$$

En primer lugar tenemos que observar que estas definiciones son correctas, en el sentido de que si se cumple

$$\frac{a}{b} = \frac{a'}{b'} \quad \text{y} \quad \frac{c}{d} = \frac{c'}{d'},$$

entonces

$$\frac{ad + bc}{bd} = \frac{a'd' + b'c'}{b'd'}, \qquad \frac{ac}{bd} = \frac{a'c'}{b'd'}.$$

Esto es una comprobación rutinaria que no ofrece ninguna dificultad. Observemos además que estas operaciones no estarían bien definidas si el anillo de partida no fuera un dominio íntegro, pues los denominadores de la suma o el producto podrían ser 0 aunque no lo fueran los de los sumandos o factores.

En segundo lugar observamos que

$$\frac{a}{1}+\frac{b}{1}=\frac{a+b}{1}, \qquad \frac{a}{1}\cdot\frac{b}{1}=\frac{ab}{1},$$

por lo que la suma y el producto de elementos del dominio íntegro de partida dan el mismo resultado cuando los calculamos con las operaciones dadas en él o si los calculamos considerándolos fracciones.

En tercer lugar, una comprobación rutinaria —pero sin dificultad alguna— muestra que las fracciones cumplen todas las propiedades de la definición de dominio íntegro. Notemos que los elementos neutros de la suma y el producto son $0/1$ y $1/1$. En general, se cumple

$$\frac{a}{b}=\frac{0}{1} \quad \text{si y sólo si} \quad a=0, \qquad \frac{a}{b}=\frac{1}{1} \quad \text{si y sólo si} \quad a=b.$$

Cuerpos de cocientes Así pues, las fracciones sobre un dominio íntegro cumplen la definición de *cuerpo* que recogemos en el resumen 1.9. A dicho cuerpo se le llama *cuerpo de cocientes* del dominio íntegro dado. En particular tenemos que los números racionales forman un cuerpo.

Observemos que un cuerpo no es más que un dominio con la condición adicional de que todo elemento no nulo a tenga un inverso a^{-1} tal que $aa^{-1}=1$.

Equivalentemente, podemos decir que un cuerpo es dominio en el que todo elemento no nulo es una unidad.

Cuando un elemento tiene inverso, éste es único, pues si hubiera otro a^*, tendríamos que

$$a^*=a^*\cdot 1=a^*\cdot a\cdot a^{-1}=1\cdot a^{-1}=a^{-1}.$$

En el caso de las fracciones hemos probado que

$$\left(\frac{a}{b}\right)^{-1}=\frac{b}{a}.$$

Observemos también que todo cuerpo es, de hecho, un dominio íntegro, pues si se cumple $ab=0$ y $a\neq 0$, entonces $b=1\cdot b=a^{-1}ab=a^{-1}\cdot 0=0$.

En general, si a es una unidad en un anillo (en particular, si a es un elemento no nulo en un cuerpo) podemos definir potencias de a con exponente entero:

$$a^n=\begin{cases}\overbrace{a\quad\cdots\quad a}^{n\text{ veces}} & \text{si } n>0,\\ 1 & \text{si } n=0,\\ \underbrace{a^{-1}\cdots a^{-1}}_{-n\text{ veces}} & \text{si } n<0.\end{cases}$$

de modo que, como es fácil comprobar, las propiedades del resumen 1.6 se siguen cumpliendo aunque los exponentes sean negativos.

Resumen 1.9: Definición de cuerpo

Un *cuerpo* es un conjunto en el que hay definidas una suma y un producto que cumplan las propiedades siguientes:

Propiedad asociativa	$(a+b)+c = a+(b+c)$
Elemento neutro	Existe un elemento 0 tal que $a+0=a$.
Elemento opuesto	Para cada a existe un $-a$ tal que $a+(-a)=0$.
Propiedad conmutativa	$a+b=b+a$
Propiedad asociativa	$(ab)c = a(bc)$
Propiedad distributiva	$a(b+c) = ab+ac$
Elemento neutro	Existe un elemento $1 \neq 0$ tal que $a \cdot 1 = a$.
Elemento inverso	Si $a \neq 0$ existe un a^{-1} tal que $a \cdot a^{-1} = 1$.
Propiedad conmutativa	$ab = ba$

Un *cuerpo ordenado* es un cuerpo con una relación de orden (Resumen 1.1) que además cumple las *relaciones de compatibilidad:*

con la suma	Si $b \leq c$ entonces $a+b \leq a+c$.
con el producto	Si $a \geq 0$ y $b \geq 0$, entonces $ab \geq 0$.

Por último, observamos que las fracciones cumplen una propiedad adicional:

Si $a/b \neq 0$, entonces $a \neq 0$, luego b/a también es una fracción, y además:

$$\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ab} = \frac{1}{1}.$$

Ordenación de cuerpos de cocientes Si partimos de un dominio íntegro ordenado, (como es el caso del anillo de los números enteros), toda fracción puede expresarse con denominador positivo, pues

$$\frac{a}{b} = \frac{-a}{-b}$$

y, si $b < 0$, entonces $-b > 0$. En el cuerpo de cocientes podemos definir una relación de orden mediante:

$$\frac{a}{b} \leq \frac{c}{d} \quad \text{si y sólo si} \quad ad \leq bc,$$

considerando únicamente fracciones con denominador positivo. Una vez más hay que comprobar (y se comprueba sin esfuerzo) que esta definición es correcta, en el sentido de que si

$$\frac{a}{b} = \frac{a'}{b'}, \qquad \frac{c}{d} = \frac{c'}{d'}, \qquad \frac{a}{b} \leq \frac{c}{d},$$

entonces

$$\frac{a'}{b'} \leq \frac{c'}{d'}.$$

A su vez observamos que

$$\frac{a}{1} \leq \frac{c}{1} \quad \text{si y sólo si} \quad a \leq c$$

así como que el cuerpo de cocientes satisface la definición de cuerpo ordenado. Esto se aplica en particular al cuerpo de los números racionales. Es evidente que

$$\left|\frac{a}{b}\right| = \frac{|a|}{|b|}.$$

Parte entera y parte fraccionaria La ordenación de los números racionales es sustancialmente distinta de la de los números enteros o naturales. No es cierto que todo número racional tenga un siguiente o un anterior. De hecho, ninguno lo tiene. Por ejemplo, es fácil ver que si $r < s$ son dos elementos de un cuerpo ordenado cualesquiera, entonces

$$r < \frac{r+s}{2} < s,$$

de modo que entre dos números racionales distintos siempre hay otro, luego ningún número racional está justo a continuación de otro. Sin embargo, sí que se cumple que todo número racional r se encuentra entre un cierto número entero y su siguiente, es decir, que puede descomponerse como $r = e + f$, donde e es un número entero y f un número racional $0 \leq f < 1$, lo que equivale a que $e \leq r < e + 1$. El número e recibe el nombre de *parte entera* de r, mientras que f es su *parte fraccionaria.*

Ahora bien, como los números racionales no son el único cuerpo ordenado que vamos a manejar, conviene justificar la existencia de la parte entera y la parte fraccionaria en un contexto más general. Lo primero es observar que los números racionales cumplen una propiedad relevante que no es necesariamente cierta en un cuerpo ordenado arbitrario:

> Un anillo ordenado es *arquimediano*[16] si para cualquier par de elementos ϵ, $M > 0$, existe un número natural n tal que $M \leq n\epsilon$.

Esto significa que si en el anillo vamos subiendo a pasos

$$0 < \epsilon < 2\epsilon < 3\epsilon < \cdots,$$

por pequeña que sea la longitud de cada paso, tras un número finito de pasos podremos superar cualquier meta M que nos marquemos.

Ahora es fácil probar:

[16] El nombre se debe a que el hecho de que el cuerpo de los números reales sea arquimediano está en la base de muchos argumentos básicos del cálculo infinitesimal de los que Arquímedes fue precursor.

1. Los números enteros forman un dominio íntegro ordenado arquimediano.

   Esto es inmediato: si ϵ, $M > 0$ son números enteros, entonces $\epsilon \geq 1$, luego $M \leq M\epsilon$ y $n = M$ cumple la definición.

2. El cuerpo de cocientes de un dominio íntegro ordenado arquimediano es un cuerpo arquimediano.

   Tomemos $\epsilon = a/b > 0$, $M = c/d > 0$, con denominadores positivos. Para que un número natural n cumpla $M \leq n\epsilon$, basta con que $c/d \leq na/b$, que a su vez equivale a $cb \leq nad$, luego basta usar la propiedad arquimediana del dominio de partida con $\epsilon' = ad > 0$ y $M' = cb > 0$.

3. En un dominio ordenado arquimediano, para cada elemento a existe un único entero e tal que $e \leq a < e + 1$.

   En efecto, por la propiedad arquimediana existe un número natural n tal que $|a| \leq n \cdot 1 = n$. Esto equivale a $-n \leq a \leq n$. Esto implica que el conjunto de los números enteros m tales que $m \leq a$ es no vacío (pues $-n$ es uno de ellos) y está acotado superiormente por n, luego podemos tomar el máximo entero e tal que $e \leq a$, y así, necesariamente $e \leq a < e + 1$ (ya que no puede ser $e + 1 \leq a$).

   Para probar la unicidad suponemos que $e' \leq a < e' + 1$, con e' entero. Si fuera $e < e'$, entonces

   $$a < e + 1 \leq e' \leq a,$$

   y tenemos una contradicción, al igual que si suponemos $e' < e$, luego tiene que ser $e = e'$.

   Si a es un elemento de un dominio ordenado arquimediano, se llama *parte entera* de a, y se representa por $E[a]$, al único número entero que cumple

   $$E[a] \leq a < E[a] + 1.$$

   El número $F[a] = a - E[a]$ se llama *parte fraccionaria de* a, y cumple que $a = E[a] + F[a]$, con $0 \leq F[a] < 1$.

Conviene tener presente que, según la definición que hemos dado:

$$E[30/7] = 4, \qquad \text{pero} \qquad E[-30/7] = -5.$$

Fracciones en cuerpos arbitrarios Conviene observar que el concepto de "fracción" que hemos usado para construir el cuerpo de cocientes a partir de un dominio íntegro dado se puede usar en cualquier cuerpo, aunque no lo hayamos definido como cuerpo de cocientes de ningún dominio íntegro. En efecto, si a y b son elementos de un cuerpo y $b \neq 0$, podemos definir la *fracción*

$$\frac{a}{b} = ab^{-1},$$

y se comprueba trivialmente que

$$\frac{a}{b} = \frac{c}{d} \quad \text{si y sólo si} \quad ad = bc,$$

así como que se cumplen las fórmulas que hemos dado para la suma y el producto de fracciones (que ahora no son definiciones, sino propiedades que podemos demostrar a partir de las propiedades de la definición de cuerpo). Más aún, en todo cuerpo ordenado las fracciones cumplen la relación que hemos tomado como definición del orden en los cuerpos de cocientes sobre dominios íntegros ordenados.

Por último conviene observar lo siguiente:

Todo cuerpo ordenado contiene a los números racionales.

En efecto, ya hemos observado que contiene a los números enteros, pero como es un cuerpo, también contiene a las fracciones de números enteros, es decir, a los números racionales, y hemos visto que la suma, el producto y el orden de las fracciones en un cuerpo ordenado coinciden con las que hemos tomado como definición en los cuerpos de cocientes, luego podemos identificar las fracciones de enteros en un cuerpo ordenado cualquiera con los números racionales.

Números combinatorios Sin más que aplicar repetidamente la propiedad distributiva y la propiedad conmutativa, en cualquier dominio podemos operar para obtener expresiones como éstas:

$$\begin{aligned}
(a+b)^0 &= 1\\
(a+b)^1 &= a+b\\
(a+b)^2 &= a^2+2ab+b^2\\
(a+b)^3 &= a^3+3a^2b+3ab^2+b^3\\
(a+b)^4 &= a^4+4a^3b+6a^2b^2+4ab^3+b^4\\
&\vdots
\end{aligned}$$

Observemos los coeficientes que van apareciendo:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 \\
\multicolumn{11}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots}
\end{array}$$

Esta configuración de números se conoce como *triángulo de Tartaglia.* Los números que aparecen en él se conocen como *números combinatorios* y se representan

así:

$$\begin{array}{ccccccccc}
 & & & & \binom{0}{0} & & & & \\
 & & & \binom{1}{0} & & \binom{1}{1} & & & \\
 & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\
 & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\
\binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{array}$$

de modo que se cumple lo que se conoce como *fórmula del binomio de Newton:*

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n.$$

Esto no sería más que una definición si no fuera porque es posible definir y calcular los números combinatorios sin necesidad de desarrollar potencias.

Para entender cómo se generan los números combinatorios pensemos en lo que supone desarrollar por la propiedad distributiva una potencia $(a+b)^n$, es decir,

$$\overbrace{(a+b)(a+b)\cdots(a+b)(a+b)}^{n \text{ veces}}.$$

Para ello tenemos que formar sumandos formados por n factores cada uno, elegidos de todas las formas posibles con uno de los términos a o b del primer factor, otro del segundo, etc. Cada uno de ellos será de la forma $a^{n-i}b^m$. Como tenemos dos posibles elecciones en el primer factor, otras dos en el segundo, etc., en total aparecerán 2^n sumandos. Los números combinatorios aparecen cuando se agrupan todos los sumandos iguales entre sí. Concretamente, $\binom{n}{m}$ es el número de sumandos $a^{n-m}b^m$ que aparecen entre el total de 2^n sumandos.

Es importante entender que no nos estamos preguntando cuántos de esos sumandos serán iguales a $a^{n-m}b^m$, porque, por ejemplo, si $a = b$, los 2^n sumandos serán iguales a a^n. Lo que nos estamos preguntando es que, si formamos todas las sucesiones posibles de n letras iguales a a o a b, en cuántas de ellas la b aparecerá exactamente m veces. Ese número es $\binom{n}{m}$. Por ejemplo, para $n = 4$, las 16 posibilidades son:

$$aaaa, aaab, aaba, aabb, abaa, abab, abba, abbb,$$

$$baaa, baab, baba, babb, bbaa, bbab, bbba, bbbb,$$

y $\binom{4}{2}$ es el número de sucesiones en las que b aparece 2 veces, es decir:

$$aabb, \quad abab, \quad abba, \quad baab, \quad baba, \quad bbaa.$$

En otros términos, las sucesiones en las que la b aparece 2 veces son las que tienen la b en las posiciones:

$$34, \quad 24, \quad 23, \quad 14 \quad 13, \quad 12.$$

Éstos son todos los subconjuntos posibles del conjunto de posiciones $1, 2, 3, 4$ con exactamente 2 elementos, y el argumento es general: el número de sucesiones en las que la b aparece m veces se corresponde con el número de conjuntos de m posiciones entre 1 y n donde podemos situar una b. En conclusión:

El número combinatorio $\binom{n}{m}$ *es el número de subconjuntos con* m *elementos que tiene un conjunto con* n *elementos.*

A su vez, esto nos proporciona una expresión explícita para los números combinatorios, que es la que suele tomarse como definición:

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

En efecto, si queremos calcular, por ejemplo, $\binom{5}{3}$, tenemos que calcular todos los subconjuntos de 3 elementos del conjunto $1, 2, 3, 4, 5$. Una forma de hacerlo es considerar todas las formas de elegir uno de estos 5 números, luego uno de los 4 restantes y luego uno de los 3 restantes, así:

123,	124,	125,	132,	134,	135,	142,	143,	145,	152,	153,	154,
213,	214,	215,	231,	234,	235,	241,	243,	245,	251,	253,	254,
312,	314,	315,	321,	324,	325,	341,	342,	345,	351,	352,	354,
412,	413,	415,	421,	423,	425,	431,	432,	435,	451,	452,	453,
512,	513,	514,	521,	523,	524,	531,	532,	534,	541,	542,	543.

El número total de casos es $5 \cdot 4 \cdot 3$ o, en general, para el cálculo de $\binom{n}{m}$, será

$$n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!}.$$

Ahora bien, los casos

$$123, \quad 132, \quad 213, \quad 231, \quad 312, \quad 321$$

corresponden todos al mismo subconjunto de $1, 2, 3, 4, 5$ e, igualmente, cada subconjunto posible se corresponde con 6 casos de los anteriores. En general, para generar todos los casos que se corresponden con un mismo subconjunto de m elementos de $1, \ldots, n$ tenemos m opciones para el primer número, $m-1$ para el segundo, etc., luego el número de casos es $m!$.

Por consiguiente, el número total de subconjuntos de m elementos de un conjunto de n elementos resulta de dividir entre $m!$ el número de casos que hemos calculado, y esto nos da la fórmula que hemos dado para $\binom{n}{m}$.

Sin embargo, hay un criterio que permite calcular los números combinatorios más fácilmente, y es que, por una parte, es claro que

$$\binom{n}{0} = \binom{n}{n} = 1$$

y por otra parte, si nos fijamos en el triángulo de Tartaglia, cada número combinatorio que no esté en un extremo es la suma de los dos que tiene por encima. Explícitamente, si $m < n$:

$$\binom{n+1}{m+1} = \binom{n}{m} + \binom{n}{m+1}.$$

Resumen 1.10: Números combinatorios

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

$$\binom{n}{m} = \binom{n}{n-m}, \quad \binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = \binom{n}{n-1} = n,$$

$$\binom{n+1}{m+1} = \binom{n}{m} + \binom{n}{m+1}.$$

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n.$$

Es fácil justificar esta relación. Basta tener en cuenta que

$$(a+b)^{n+1} = (a+b)(a+b)^n = a(a+b)^n + b(a+b)^n.$$

El miembro derecho es la suma:

$$\begin{aligned}&\tbinom{n}{0}a^{n+1} + \tbinom{n}{1}a^n b + \tbinom{n}{2}a^{n-1}b^2 + \cdots + \tbinom{n}{n-1}a^2b^{n-1} + \tbinom{n}{n}ab^n\\ &\quad + \tbinom{n}{0}a^n b + \tbinom{n}{1}a^{n-1}b^2 + \cdots + \tbinom{n}{n-2}a^2b^{n-1} + \tbinom{n}{n-1}ab^n + \tbinom{n}{n}b^{n+1},\end{aligned}$$

de donde concluimos que

$$\binom{n}{0} + \binom{n}{1} = \binom{n+1}{1}, \quad \binom{n}{1} + \binom{n}{2} = \binom{n+1}{2}, \quad \text{etc.}$$

El resumen 1.10 contiene los hechos que hemos probado sobre los números combinatorios junto con algunas propiedades obvias más.

Veamos una aplicación de la fórmula del binomio:

> *En un cuerpo ordenado, si* $0 < a < b$ *y* n *es un número natural no nulo, entonces* $a^n < b^n$*. En particular, si* $a, b > 0$ *cumplen* $a^n = b^n$*, entonces* $a = b$*.*

En efecto, basta observar que $b^n = (a + (b-a))^n$ y, al desarrollar por la fórmula del binomio, queda $b^n = a^n + \cdots$, donde todos los términos siguientes son > 0, luego $a^n < b^n$.

Una propiedad similar, pero que se prueba trivialmente sin necesidad de la fórmula del binomio es que si $a > 1$ y $m < n$ son números enteros, entonces $a^m < a^n$. (Notemos que esto equivale a que $a^{n-m} > 1$.) En particular, si $a^m = a^n$, necesariamente $m = n$.

Fracciones egipcias Así como los antiguos griegos trabajaban eficientemente con números racionales (positivos) arbitrarios, no ocurría lo mismo con otras culturas. Por ejemplo, los egipcios consideraban únicamente fracciones con numerador 1, y expresaban cualquier otro número racional como suma de estas fracciones unitarias. Hasta aquí no parece que la diferencia sea sustancial, pues sólo supone escribir $1/6 + 1/6 + 1/6 + 1/6 + 1/6$ en lugar de $5/6$. Es un poco más farragoso, pero equivalente. Sin embargo, el caso es que los egipcios se autoimponían la absurda regla de no repetir denominadores, de modo que para referirse a $5/6$ escribían $1/2 + 1/3$.

Por ello, las expresiones de números racionales como sumas de fracciones unitarias con denominadores distintos se conocen como *fracciones egipcias.*

Las fracciones egipcias plantean varios problemas de interés en teoría de números. Aquí vamos a considerar únicamente el más básico:

Todo número racional puede expresarse como fracción egipcia.

En efecto, no perdemos generalidad si consideramos números $0 < a/b < 1$, pues todo número racional puede expresarse como $m/1 + a/b$, con $0 < a/b < 1$. Dividimos $b = ac + r$, con $0 \le r < a$. Si $r = 0$ es que $a/b = 1/c$, y ya tenemos la expresión requerida. En caso contrario $b < ac + a = a(c + 1)$, luego

$$\frac{1}{c} < \frac{a}{b}, \qquad \frac{a}{b} - \frac{1}{c+1} = \frac{ac + a - b}{b(c+1)} = \frac{a - r}{b(c+1)}.$$

Por lo tanto, podemos descomponer

$$\frac{a}{b} = \frac{1}{c+1} + \frac{a-r}{b(c+1)}$$

y la tercera fracción tiene denominador mayor que las dos primeras y numerador menor que la primera, luego al cabo de un número finito de pasos el numerador del resto será 1 y tendremos una descomposición en fracción egipcia con denominadores estrictamente crecientes, ya que al repetir el proceso dividiendo $b(c + 1) = (a - r)c' + r'$, necesariamente $c' > c$, ya que en caso contrario tendríamos $b(c + 1) < (a - r)c + a - r = (a - r)(c + 1)$ y $b < a - r < a$. Este procedimiento se debe a Fibonacci, pero existen otros y no todos dan lugar a la misma representación. Por ejemplo, si la aplicamos a $5/121$ obtenemos

$$\frac{5}{121} = \frac{1}{25} + \frac{1}{757} + \frac{1}{763\,309} + \frac{1}{873\,960\,180\,913},$$

cuando una alternativa más simple es

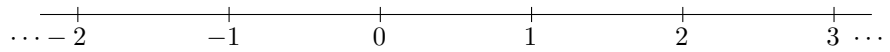
$$\frac{5}{121} = \frac{1}{33} + \frac{1}{121} + \frac{1}{363}.$$

■

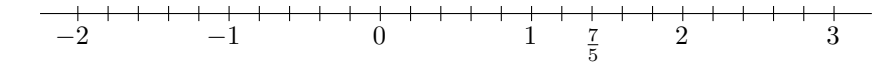
1.5 Los números reales

Los números naturales sirven para medir magnitudes discretas, es decir, propiedades cuantificables en términos de unidades indivisibles, mientras que los números racionales son un primer paso hacia el tratamiento matemático de las "magnitudes continuas", concepto que aún no estamos en condiciones de definir con precisión, pero que en particular son "infinitamente divisibles", como es el caso de las longitudes, las áreas, etc., en las que no cabe hablar de una mínima longitud indivisible, ni de una mínima área indivisible, etc.

Para precisar esta idea, consideremos una recta y fijemos en ella dos puntos arbitrariamente, a los que asignamos los números 0 y 1. Tomamos la distancia entre ambos puntos como unidad de longitud y asignamos el número natural n al punto que se encuentra a n unidades de distancia del 0 en la semirrecta que contiene al 1, mientras que asignamos el número $-n$ al punto que se encuentra a n unidades de distancia en la semirrecta opuesta, así:

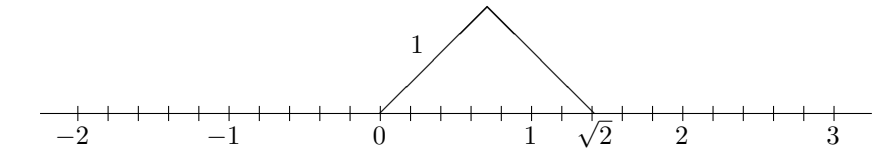


Podemos expresar esto diciendo que hemos asignado una "coordenada" entera a ciertos puntos de la recta. Más aún, podemos asignar un punto de la recta a cada número racional. Por ejemplo, para "situar" el número 7/5 dividimos los intervalos unitarios en 5 partes iguales y tomamos 7 de ellas:



A los números negativos como $-7/5$ les asignamos puntos con el mismo criterio, pero en la semirrecta opuesta. Es fácil convencerse de que el punto obtenido depende realmente del número racional y no de la representación como fracción que usamos para determinarlo, así como que a cada número racional le corresponde un punto distinto en la recta.

Sin embargo, la geometría nos enseña que hay puntos en la recta que no tienen asignado ningún número racional. No estamos ahora en condiciones de justificarlo, pero sin duda el lector conocerá el teorema de Pitágoras, que implica que la longitud h de la hipotenusa de un triángulo rectángulo cuyos catetos tengan longitud 1 debe cumplir $h^2 = 1^2 + 1^2 = 2$, es decir, que si el punto marcado como $\sqrt{2}$ en la figura siguiente:



tiene que tener asociado un número racional h, éste debe cumplir $h^2 = 2$. Sin embargo, veremos más adelante (véase el final de la sección 2.1) que ningún número racional tiene esta propiedad, luego si el punto construido en la figura

Resumen 1.11: Los números reales

Los *números reales* forman un cuerpo ordenado completo, donde la *completitud* es la propiedad siguiente:

> Si dividimos la totalidad de los números reales en dos conjuntos no vacíos A y B de modo que cada elemento de A es menor que cada elemento de B, entonces existe un número real que, o bien es el máximo de los elementos de A, o bien es el mínimo de los elementos de B (según si está en A o en B).

En particular, los números reales son un cuerpo ordenado arquimediano.

tiene que tener asociada una coordenada numérica, ésta tiene que ser un "número irracional".

Es aquí donde entran en escena los números reales. El cuerpo $\mathbb{R}$ de los números reales es un cuerpo ordenado arquimediano, es decir, un cuerpo que cumple las mismas propiedades algebraicas básicas que el cuerpo $\mathbb{Q}$ de los números racionales, pero que contiene una raíz cuadrada de 2 y, en general, todos los números irracionales necesarios para que pueda establecerse una correspondencia biunívoca natural que nos permite identificarlos con los puntos de una recta.

Vamos a ver que los números reales están completamente caracterizados por las propiedades de "cuerpo ordenado" y una propiedad adicional que incluimos en el Resumen 1.11. Esta propiedad de completitud puede expresarse de muchas formas equivalentes, de entre las cuales hemos escogido la más clara desde un punto de vista geométrico, que afirma que si podemos dividir la recta en dos mitades A y B, tiene que haber un punto que constituya el final de A y el principio de B:

Lo contrario significaría que la recta contiene "un agujero microscópico" entre A y B. La completitud expresa que la recta "no tiene agujeros".

Como ejemplo del uso práctico de la propiedad de completitud vamos a demostrar que, tal y como se afirma en el Resumen 1.11:

Los números reales son un cuerpo ordenado arquimediano.

Esto supone probar que si $M > 0$ es un número real, existe un número natural n tal que $M < n$, pues, si se cumple esto, basta aplicarlo a M/ϵ para tener el caso general. Supongamos, por el contrario, que existe un número $M > 0$ tal que $n \leq M$ para todo número natural n. Entonces podemos definir A como el conjunto de todos los números reales que son menores que algún número natural y B como el conjunto de los números reales que son mayores que todos los números naturales.

Claramente A no es vacío, y estamos suponiendo que B tampoco lo es. También es obvio que si $x \in A$, $y \in B$, entonces, existe un número natural n tal que $x < n < y$, luego A y B cumplen las hipótesis de la propiedad de completitud de $\mathbb{R}$, lo que nos permite concluir que existe un número real α que, o bien es el máximo de A, o bien es el mínimo de B. Lo primero es imposible, pues supondría en particular que $\alpha \in A$, luego existiría un número natural n tal que $\alpha < n$, pero entonces $\alpha + 1 < n + 1$, luego $\alpha < \alpha + 1 \in A$, en contra de que α era el máximo elemento de A. Por lo tanto, α es el mínimo elemento de B, pero entonces $\alpha - 1$ no puede estar en B, luego está en A, luego existe un número natural tal que $\alpha - 1 < n$, luego $\alpha < n + 1$, lo cual implica que $\alpha + 1$ está en A, cuando debía estar en B, y tenemos una contradicción. ∎

A su vez, la propiedad arquimediana tiene como consecuencia que los números racionales e irracionales están muy "mezclados":

> *Entre dos números reales cualesquiera hay siempre números racionales e irracionales.*

En efecto, sean $\alpha < \beta$ dos números reales y sea $\gamma > 0$ un número racional (resp. irracional arbitrario). En principio, todavía no hemos probado que existen números irracionales, pero esto será pronto evidente. Por la propiedad arquimediana, existe un número natural n tal que $\gamma < n(\beta - \alpha)$. En un cuerpo arquimediano está definida la parte entera, luego podemos tomar $m = E[n\alpha/\gamma]$, de modo que

$$m \leq \frac{n\alpha}{\gamma} < m + 1,$$

lo que equivale a

$$\frac{m\gamma}{n} \leq \alpha < \frac{m\gamma}{n} + \frac{\gamma}{n} < \alpha + \beta - \alpha = \beta.$$

El número $\xi = (m + 1)\gamma/n$ es claramente racional o irracional según lo sea γ (pues si ξ es racional, entonces $\gamma = n\xi/(m + 1)$ también lo es), y está situado entre α y β. ∎

Veamos una segunda aplicación de la propiedad de completitud:

> *Si $n \geq 1$ es un número natural y $a \geq 0$ es un número real, existe un único número real $b \geq 0$ tal que $b^n = a$.*

Demostración: El resultado es obviamente cierto para $a = 0$, así que podemos suponer $a > 0$. Llamamos A al conjunto de todos los números reales $x \leq 0$ y los que cumplen $x > 0$ y $x^n < a$. Por otra parte, sea B el conjunto de los números reales $x > 0$ que cumplen $x^n \geq a$.

Es claro entonces que todo número real está en A o en B, así como que cada elemento de A es menor que cada elemento de B (pues si x está en A e y está en B, o bien $x \leq 0 < y$, o bien $x > 0$ y $x^n < a \leq y^n$, de donde $x < y$, pues si fuera $0 < y \leq x$, tendríamos que $y^n \leq x^n$).

Ejercicio: Probar que A y B no son vacíos.

La completitud de $\mathbb{R}$ implica que existe un b que, o bien es el máximo elemento de A, o bien es el mínimo elemento de B. Vamos a probar que $b^n = a$.

En primer lugar observamos que no puede suceder que b sea el máximo de A, pues entonces $b^n < a$, pero, para cada número real $0 < \epsilon < 1$, tenemos que

$$(b+\epsilon)^n - b^n = \epsilon(\binom{n}{1}b^{n-1} + \binom{n}{2}b^{n-2}\epsilon + \cdots + \binom{n}{n}\epsilon^{n-1}) \leq \epsilon M,$$

donde $M = \binom{n}{1}b^{n-1} + \binom{n}{2}b^{n-2} + \cdots + \binom{n}{n}$.

Tomando $0 < \epsilon < (a-b^n)/M$, tenemos que $(b+\epsilon)^n - b^n < a - b^n$, luego $(b+\epsilon)^n < a$, luego $b+\epsilon$ está en A y $b < b+\epsilon$, en contra de que b sea el máximo de A.

Por lo tanto, b tiene que ser el mínimo de B y, en particular, $a \leq b^n$. Supongamos que $a < b^n$ y llegaremos a una contradicción. Tomamos de nuevo $0 < \epsilon < 1$, $\epsilon < b$, de modo que

$$b^n - (b-\epsilon)^n = \epsilon(-\binom{n}{1}b^{n-1} + \binom{n}{2}b^{n-2}\epsilon - \cdots + (-1)^n\binom{n}{n}\epsilon^{n-1}) \leq \epsilon M,$$

donde M depende de b, pero no de ϵ. Ahora tomamos $0 < \epsilon < (b^n - a)/M$, de modo que

$$b^n - (b-\epsilon)^n \leq b^n - a,$$

y así $a < (b-\epsilon)^n$, luego $b-\epsilon$ está en B y eso contradice que b sea el mínimo de B. ∎

El número b se llama *raíz n-sima* (positiva) de a, y se representa por $\sqrt[n]{a}$. Cuando $n = 2$ es costumbre omitir el índice y escribir simplemente $\sqrt{a}$.

De la propia definición se siguen inmediatamente las propiedades siguientes:

$$\sqrt[n]{ab} = \sqrt[n]{a}\,\sqrt[n]{b}, \qquad \sqrt[n]{1/a} = 1/\sqrt[n]{a}, \quad (\sqrt[n]{a})^m = \sqrt[n]{a^m}.$$

Desarrollos decimales Observemos que en ningún momento hemos precisado qué debemos entender exactamente por "número real". Sólo hemos dicho que los números reales forman un cuerpo ordenado completo y hemos extraído varias consecuencias de estas propiedades, pero podría haber muchos cuerpos ordenados completos con propiedades distintas, de modo que, por lo que sabemos hasta ahora, no podríamos saber cuáles son válidas para los números reales y cuáles no. Pero vamos a ver que no se da el caso, sino que el hecho de ser un cuerpo ordenado completo caracteriza completamente a los números reales. Concretamente, vamos a ver que, al igual que un número natural como 45 020 está completamente determinado por sus cifras decimales (o binarias, o en cualquier base) lo mismo sucede con los números reales, con la diferencia de que, en general, para determinar un número real hacen falta infinitas cifras.

Llamamos *números decimales exactos* a los números racionales de la forma $a/10^n$, donde a es un número natural. En realidad podemos sustituir el 10 por cualquier número natural $c \geq 2$, pero por simplicidad vamos a trabajar con $c = 10$, que es lo más habitual.

La notación usual para los decimales exactos es la que ilustra el ejemplo siguiente:

$$\frac{14142}{10^4} = 1.4142.$$

Así, el punto decimal indica que el número que resulta de eliminarlo debe dividirse entre 10 elevado al número de dígitos que quedan a su izquierda. Alternativamente:

$$1.4142 = 1 + \frac{4}{10} + \frac{1}{10^2} + \frac{4}{10^3} + \frac{2}{10^4}.$$

Esta expresión muestra que, en general, podemos descomponer

$$25.23452 = 25 + 0.23452,$$

y es claro que un decimal exacto que empiece por "0." está siempre entre 0 y 1. Por ejemplo,

$$0.23452 = \frac{23452}{100\,000} < 1.$$

Por lo tanto, si $\alpha = 25.23452$, tenemos que 25 es lo que hemos llamado la parte entera de α, mientras que 0.23452 es su parte fraccionaria.

Si tomamos un número real arbitrario $\alpha > 0$, por ejemplo, $\alpha = \sqrt{2}$, podemos situarlo entre dos números naturales: $c_0 \leq \alpha \leq c_0+1$. En nuestro ejemplo, como $1^2 < 2 < 2^2$, resulta que

$$1 < \sqrt{2} < 2.$$

A su vez, podemos dividir el intervalo comprendido entre c_0 y $c_0 + 1$ en 10 subintervalos:

$$1 < 1.1 < 1.2 < 1.3 < 1.4 < 1.5 < 1.6 < 1.7 < .18 < 1.9 < 2$$

El número α tendrá que estar comprendido entre dos de estos números:

$$c_0 + \frac{c_1}{10} \leq \alpha \leq c_0 + \frac{c_1 + 1}{10},$$

donde $0 \leq c_1 \leq 9$. En nuestro ejemplo, calculando

x	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
x^2	1	1.21	1.44	1.69	1.96	2.25	2.56	2.89	3.24	3.61	4

vemos que $1.4^2 < 2 < 1.5^2$, por lo que $1.4 < \sqrt{2} < 1.5$.

A su vez, dividimos el intervalo comprendido entre $c_0+c_1/10$ y $c_1+(c_1+1)/10$ en diez subintervalos, y elegimos el que cumple

$$c_0 + \frac{c_1}{10} + \frac{c_2}{100} \leq \alpha \leq c_0 + \frac{c_1}{10} + \frac{c_2 + 1}{100}.$$

En el caso de $\alpha = \sqrt{2}$, es fácil ir comprobando que

$$1 < \sqrt{2} < 2$$

$$1.4 < \sqrt{2} < 1.5$$

$$1.41 < \sqrt{2} < 42$$

$$1.414 < \sqrt{2} < 1.415$$

$$1.4142 < \sqrt{2} < 1.4143$$

$$1.41421 < \sqrt{2} < 1.41422$$

de modo que 1 se diferencia de $\sqrt{2}$ en menos de una unidad, 1.4 se diferencia de $\sqrt{2}$ en menos de una décima, 1.41 en menos de una centésima, 1.414 en menos de una milésima, etc. Así podemos obtener un desarrollo decimal infinito:

$$\sqrt{2} = 1.41421356237309504880168872420969807856967187537694\ldots$$

que determina a $\sqrt{2}$ en el sentido de que al truncar la expresión por su n-sima cifra decimal obtenemos una aproximación racional de $\sqrt{2}$ con un error menor que 10^{-n}. Explícitamente:

La igualdad

$$\alpha = c_0.c_1c_2c_3c_4c_5\ldots$$

donde c_0 es un número natural y $0 \leq c_i \leq 9$, para $i \geq 1$, significa que el número real α es el único que cumple

$$\alpha_n \leq \alpha \leq \alpha_n + \frac{1}{10^n},$$

donde $\alpha_n = c_0.c_1\cdots c_n$ es el decimal exacto determinado por las n primeras cifras decimales de la sucesión.

Si un número real $\alpha > 0$ cumple esto, ciertamente es único, pues si hubiera otro β que cumpliera lo mismo, por ejemplo con $\alpha < \beta$, entonces habría un n tal que

$$\frac{1}{\beta - \alpha} < 10^n,$$

pues todo número real es menor que un número natural, y todo número natural es menor que una potencia de 10. Pero entonces

$$\alpha_n \leq \alpha < \beta \leq \alpha_n + 10^{-n},$$

pero esto implica que $\beta - \alpha \leq 10^{-n}$, en contradicción con la elección de n.

El proceso con el que hemos obtenido el desarrollo decimal de $\sqrt{2}$ (es decir, el proceso de ir subdividiendo intervalos en 10 subintervalos determinados por decimales exactos y tomando uno que contenga al número) puede realizarse igualmente con cualquier número real positivo, lo que prueba que todo número real positivo puede determinarse mediante una sucesión de decimales.

Recíprocamente, toda sucesión de decimales

$$c_0.c_1c_2c_3c_4c_5\ldots$$

determina un (único) número real. En efecto, la sucesión α_n de decimales exactos que resulta de tomar únicamente los n primeros decimales es creciente:

$$\alpha_0 \leq \alpha_1 \leq \alpha_2 \leq \alpha_3 \leq \alpha_4 \leq \alpha_5 \leq \alpha_6 \leq \alpha_7 \leq \cdots$$

Podemos dividir $\mathbb{R}$ en dos conjuntos: un conjunto A formado por todos los números reales β que cumplen $\beta < \alpha_n$ para algún n, y otro B formado por los números reales β que cumplen $\alpha_n \leq \beta$ para todo n. Obviamente, cada elemento de A es menor que cada elemento de B, y el conjunto A no tiene máximo, pues si $\beta < \alpha_n$, podemos tomar $\beta < \beta' < \alpha_n$, así que β' también está en A y es mayor que β. La completitud de $\mathbb{R}$ implica entonces que B tiene un mínimo elemento α. Sólo tenemos que probar que

$$\alpha_n < \alpha \leq \alpha_n + \frac{1}{10^n},$$

para todo n. En efecto, la primera desigualdad se cumple porque α está en B. Para probar la segunda basta ver que $\alpha_m < \alpha_n + 10^{-n}$, para todo m. Esto es obvio si $m \leq n$ y, en caso contrario,

$$\alpha_m = \alpha_n + \frac{c_{n+1}}{10^{n+1}} + \cdots + \frac{c_m}{10^m} \leq \alpha_n + \frac{9}{10^{n+1}}\left(1 + \frac{1}{10} + \cdots + \left(\frac{1}{10}\right)^{m-n-1}\right)$$

$$= \alpha_n + \frac{9}{10^{n+1}}\frac{(1/10)^{m-n}-1}{(1/10)-1} < \alpha_n + \frac{9}{10^{n+1}}\frac{1}{1-1/10} = \alpha_n + \frac{1}{10^n}.$$

Sin embargo, a veces hay que tener presente que un mismo número real puede admitir dos desarrollos decimales distintos. El caso más simple es

$$1 = 1.00000\ldots = 0.99999\ldots$$

En efecto, basta tener en cuenta que en ambos casos se cumplen las relaciones $\alpha_n \leq 1 \leq \alpha_n + 10^{-n}$. Explícitamente:

1	$\leq 1 \leq 2$	0	$\leq 1 \leq 1$
1.0	$\leq 1 \leq 1.1$	0.9	$\leq 1 \leq 1$
1.00	$\leq 1 \leq 1.01$	0.99	$\leq 1 \leq 1$
1.000	$\leq 1 \leq 1.001$	0.999	$\leq 1 \leq 1$
1.0000	$\leq 1 \leq 1.0001$	0.9999	$\leq 1 \leq 1$
1.00000	$\leq 1 \leq 1.00001$	0.99999	$\leq 1 \leq 1$
	$\ldots$		$\ldots$

En otras palabras, el 1 es un decimal exacto que puede aproximarse por sí mismo, pero también por la sucesión de decimales exactos 0.9, 0.99, 0.999, etc. Lo mismo vale para cualquier otro decimal exacto:

$$143.265599999\ldots = 143.2655 + 10^{-4}\cdot 0.9999\ldots = 143.2655 + 0.0001 = 143.2656.$$

En general, una expresión decimal finalmente igual a 9 aproxima al decimal exacto que resulta de eliminar los nueves y sumar 1 a la cifra decimal precedente.

No es difícil convencerse de que ésta es la única excepción a la unicidad de los desarrollos decimales. En efecto, si α es un decimal exacto, por ejemplo, $\alpha = 143.2656$, sus dos únicos desarrollos decimales son:

143	$< \alpha < 144$	143	$< \alpha < 144$
143.2	$< \alpha < 143.3$	143.2	$< \alpha < 143.3$
143.26	$< \alpha < 143.27$	143.26	$< \alpha < 143.27$
143.265	$< \alpha < 143.266$	143.265	$< \alpha < 143.266$
143.2655	$< \alpha \leq 143.2656$	143.2656	$\leq \alpha < 143.2657$
143.26559	$< \alpha \leq 143.26560$	143.26560	$\leq \alpha < 143.26561$
143.265599	$< \alpha \leq 143.265600$	143.265600	$\leq \alpha < 143.265601$
143.2655999	$< \alpha \leq 143.2656000$	143.2656000	$\leq \alpha < 143.2656001$
	$\ldots$		$\ldots$

Partimos de que $143 < \alpha < 144$ y, al ir dividiendo cada intervalo en 10 subintervalos, α se encuentra en uno solo de ellos hasta que llegamos al intervalo

$$143.265 < \alpha < 143.266.$$

En este punto tenemos que considerar los 10 subintervalos

$$143.2650 < 143.2651 < 143.2652 < 143.2653 < 143.2654 < 143.2655$$

$$< \alpha = 143.2656 < 143.2657 < 143.2658 < 143.2659 < 143.2660,$$

y sucede que α es el extremo superior de uno de ellos y el extremo inferior del siguiente, por lo que podemos prolongar la sucesión de intervalos de dos formas distintas (como se ve en la línea destacada de los dos desarrollos precedentes). Ahora bien, si optamos por el intervalo cuyo extremo derecho es α, en todas las subdivisiones siguientes α será el extremo derecho del último subintervalo, luego el desarrollo decimal continuará necesariamente con una sucesión de nueves, mientras que si optamos por el intervalo cuyo extremo izquierdo es α, en todas las subdivisiones siguientes α será el extremo izquierdo del primero de los subintervalos, luego la sucesión decimal continuará necesariamente con ceros.

Por lo tanto, concluimos que un decimal exacto tiene exactamente dos sucesiones decimales. Por otro lado, para que esto pueda suceder, es decir, para que α sea el extremo común de dos subintervalos, es necesario que α sea un decimal exacto (pues los extremos de los intervalos siempre lo son), de modo que los números que no son decimales exactos sólo admiten un único desarrollo decimal.

Todo esto se aplica a números reales positivos. Es fácil ver que 0 sólo admite el desarrollo decimal $0 = 0.0000\ldots$ y cada número negativo se expresa en términos del desarrollo decimal de su opuesto. Por ejemplo,

$$-\frac{1}{3} = -0.33333\ldots$$

Así pues, ahora ya tenemos una idea clara de qué son los números reales más allá de la mera descripción algebraica de que son un cuerpo ordenado completo: Un número real es un objeto determinado por un signo, un número natural y una sucesión de infinitas cifras decimales (teniendo en cuenta que en un caso excepcional perfectamente delimitado dos sucesiones distintas de decimales pueden corresponder al mismo número real).

Preguntarse qué es un número real más allá de esta descripción es un sinsentido, como lo sería preguntar qué es el número natural 7 más allá del hecho de que es el número que va después del 6 y antes del 8.

Decimales finalmente periódicos Consideremos el desarrollo decimal de un número racional, como

$$\frac{67}{44} = 1.5227\ldots$$

Esto significa que

$$1.5227 \leq \frac{67}{44} \leq 1.5228$$

o, equivalentemente,

$$15\,227 \leq \frac{67}{44} \cdot 10^4 \leq 15\,227 + 1$$

lo que a su vez equivale a que

$$670\,000 = 44 \cdot 15\,227 + r, \qquad 0 \leq r \leq 44.$$

Vemos así que, en general, para obtener las n primeras cifras decimales de un número racional a/b, basta dividir euclídeamente $a \cdot 10^n$ entre b. En la práctica, esto significa que las cifras decimales se obtienen sin más que prolongar la división euclídea añadiendo ceros al dividendo:

$$\begin{array}{r|l} 67.0000 & 44 \\ \cline{2-2} 23\,0 & 1.5227 \\ 1\,00 & \\ 120 & \\ 320 & \\ 12 & \end{array}$$

En realidad ya no necesitamos prolongar más el cálculo, pues, como se ha repetido el resto 12, ya sabemos que la próxima cifra decimal volverá a ser 2, que dará lugar a un resto de 32, con lo que la cifra siguiente será otra vez 7, que volverá a dar resto 12, etc.

Así pues, ahora sabemos que el desarrollo decimal

$$\frac{67}{44} = 1.52272727\ldots$$

consta de la parte entera (1) un anteperiodo decimal (52) y un periodo (27).

Esto no es casual, sino que, dado que los restos sólo pueden variar entre 0 y 44, era necesario que en algún momento dado (a lo sumo, al cabo de 45 pasos), un resto apareciera repetido, y a partir de ese momento las cifras decimales tenían que repetirse cíclicamente.

En resumen, esto prueba que los desarrollos decimales de los números racionales son finalmente periódicos (entendiendo que los decimales exactos tienen periodo 0).

Recíprocamente, todo número decimal finalmente periódico corresponde a un número racional. Por ejemplo, consideremos el número

$$\alpha = 12.345678967896789\ldots$$

que tiene parte entera 12, anteperiodo 345 y periodo 6789.

Para expresarlo como fracción calculamos

$$10^7 \cdot \alpha = 123\,456\,789.67896789\ldots = 123\,456\,789 + 0.67896789\ldots$$

(donde el exponente 7 es la suma de las longitudes del anteperiodo y el periodo).

$$10^3\alpha = 12\,345.678967896789\ldots = 12\,345 + 0.678967896789$$

(donde el exponente 3 es la longitud del anteperiodo) y así

$$9\,999\,000\alpha = 10^7 \cdot \alpha - 10^3\alpha = 123\,456\,789 - 12345,$$

luego

$$\alpha = \frac{123\,456\,789 - 12\,345}{9\,999\,000} = \frac{10\,287\,037}{832\,500}$$

En general, hemos obtenido lo siguiente:

> *Los números decimales finalmente periódicos son precisamente los números racionales.*
>
> *Para calcular el desarrollo decimal de un número racional basta prolongar la división euclídea añadiendo ceros al dividendo.*
>
> *Para expresar como fracción un número decimal finalmente periódico restamos el número natural formado con las cifras de su parte entera, más el anteperiodo, más el periodo menos el número formado por la parte entera más el anteperiodo, y el resultado lo dividimos entre el número formado por tantos nueves como la longitud del periodo seguido de tantos ceros como la longitud del anteperiodo.*

La suma y el producto de números reales Ya hemos señalado que los números reales no son más que todas las expresiones de la forma

$$\pm 32.252023\ldots$$

con las precisiones necesarias sobre la falta de unicidad de los desarrollos decimales. Sin embargo, con esto no podemos decir que "sabemos" qué son los números

reales, pues no hemos dicho nada sobre cómo se suman o se multiplican. Al haber partido de que los números reales forman un cuerpo ordenado completo, estamos admitiendo que hay definida en ellos una suma y un producto, pero ¿no podría haber distintas formas de sumar y multiplicar números decimales, que tuvieran propiedades distintas, pero que dieran lugar en cualquier caso a una estructura de cuerpo ordenado completo? Es fácil ver que la respuesta es negativa:

> *Si α y β son números reales, entonces $\alpha+\beta$ es el único número real que cumple*
>
> $$r+s<\alpha+\beta<r'+s',$$
>
> *para todos los números racionales $r<\alpha<r'$, $s<\beta<s'$.*

Por lo tanto, la suma de números racionales determina completamente la suma de números reales. No hay más que una suma posible.

En efecto, el hecho de que entre dos números reales haya siempre números racionales implica que siempre podemos encontrar números racionales que cumplan $r<\alpha<r'$, $s<\beta<s'$, y las propiedades de cuerpo ordenado implican que entonces $r+r'<\alpha+\alpha'<s+s'$. Sólo tenemos que probar que hay un único número γ que cumple esto. Pero si hubiera dos números $\gamma<\gamma'$ que cumplieran

$$r+s<\gamma<\gamma'<r'+s'$$

para todas las elecciones posibles de r,s,r',s', consideramos $\epsilon=\gamma'-\gamma$ y tomamos números racionales que cumplan

$$\alpha-\frac{\epsilon}{4}<r<\alpha<r'<\alpha+\frac{\epsilon}{4},\qquad \beta-\frac{\epsilon}{4}<s<\beta<s'<\beta+\frac{\epsilon}{4}.$$

Entonces

$$\epsilon=\gamma'-\gamma<r'+s'-r-s<\alpha+\beta+\frac{\epsilon}{2}-\alpha-\beta+\frac{\epsilon}{2}=\epsilon,$$

y tenemmos una contradicción. ∎

Similarmente:

> *Si α,β son dos números reales positivos, entonces $\alpha\beta$ es el único número real que cumple*
>
> $$rs<\alpha\beta<r's',$$
>
> *para todos los números racionales $0<r<\alpha<r'$, $0<s<\beta<s'$.*

Así, el producto de números reales positivos está completamente determinado por el producto de números racionales positivos (y el producto de números reales arbitrarios está determinado por las reglas algebraicas para los signos: más por menos es menos, etc.).

En efecto, como en el caso de la suma, lo único que hay que comprobar es la unicidad. Supongamos que existen dos números que cumplen

$$rs < \gamma < \gamma' < r's',$$

para todas las elecciones posibles de r, s, r', s' y consideramos

$$\epsilon = \gamma' - \gamma < r's' - rs = r's' - r's + r's - rs = r'(s' - s) + s(r' - r) < M(s' - s + r' - r),$$

donde M es cualquier número real mayor que α y β, suponiendo que tomamos $\alpha < r' < M$.

Más concretamente, elegimos números racionales que cumplan

$$\alpha - \frac{\epsilon}{4M} < r < \alpha < r' < \alpha + \frac{\epsilon}{4M}, \qquad \beta - \frac{\epsilon}{4M} < s < \beta < s' < \beta + \frac{\epsilon}{4M},$$

así como $r' < M$, y entonces queda que

$$\epsilon < M\left(\frac{\epsilon}{2M} + \frac{\epsilon}{2M}\right) = \epsilon,$$

y de nuevo tenemos una contradicción. ∎

Con esto ya tenemos justificado que los números reales (junto con su relación de orden, su suma y su producto) son únicos, en el mismo sentido en que son únicos los números naturales, enteros o racionales. Cualquiera que hable de números reales (entendidos como un cuerpo ordenado completo) está hablando de sucesiones de decimales con una suma y un producto determinados por los dos resultados precedentes.[17]

Cálculo de raíces cuadradas Antiguamente se enseñaba en las escuelas un algoritmo para calcular raíces cuadradas. Actualmente carece de interés práctico, ya que es laborioso y una calculadora proporciona inmediatamente el resultado. No obstante, tiene interés histórico, pues explica cómo matemáticos de siglos pasados pudieron llevar a cabo algunos cálculos que llevaremos a cabo en este libro.

Dado un número natural N, vamos a calcular la parte entera de $\sqrt{N}$. Más precisamente, vamos a calcular los números naturales n y r que cumplen

$$n^2 \le n^2 + r = N < (n+1)^2.$$

Para $N = 7\,446\,968$, el cálculo puede disponerse en la forma siguiente:

[17] En realidad hay una sutileza que se nos está escapando aquí, y es que estamos diciendo que los números reales son los objetos determinados por *todas* las sucesiones de decimales posibles, y la lógica matemática nos enseña que ese "todas" no está bien definido, pero esto es algo que trasciende al contenido de este libro y que será irrelevante en todo lo que sigue.

$$\begin{array}{r|l}
\sqrt{\;7\,44\,69\,68} & 2728 \\
\cline{2-2}
\;\;4 & 47 \\
\cline{2-2}
\;34\,4 & 542 \\
\cline{2-2}
\;32\,9 & 5448 \\
\cline{2-2}
\;\;1\,56\,9 & \\
\;\;1\,08\,4 & \\
\cline{1-1}
\;\;\;48\,568 & \\
\;\;\;43\,584 & \\
\cline{1-1}
\;\;\;\;4\,984 &
\end{array}$$

La conclusión es que

$$2\,728^2 < 2\,728^2 + 4\,984 = 7\,446\,968 < 2\,729^2.$$

El cálculo se realiza como sigue:

1. Dividimos las cifras del radicando N en grupos de dos, empezando por la derecha. En este caso quedan 4 grupos (admitiendo que el grupo de la derecha pueda tener una sola, como en este caso).

   Esto nos dice que $10^{2\cdot 3} \le N < 10^{2\cdot 4}$, por lo que el valor de n que estamos buscando cumplirá $10^3 \le n < 10^4$, es decir, que n será un número de 4 cifras.

2. Calculamos la mayor cifra c_0 tal que c_0^2 no excede al grupo de dos cifras del radicando situado más a la izquierda. En este caso $c_0 = 2$ cumple que $2^2 \le 7$.

3. Escribimos $c_0 = 2$ en la caja que contendrá el resultado, restamos $7 - c_0^2 = 3$ y bajamos las dos cifras siguientes del radicando (44).

4. Para calcular la cifra siguiente multiplicamos c_0 por 2 y ponemos el resultado debajo:

$$\begin{array}{r|l}
\sqrt{\;7\,44\,69\,68} & 2 \\
\cline{2-2}
\;\;4 & 4 \\
\cline{2-2}
\;34\,4 &
\end{array}$$

5. Ahora buscamos la mayor cifra c_1 que cumple $4c_1 \times c_1 \le 344$ (donde $4c_1$ no es un producto, sino $40 + c_1$). En este caso es $c_1 = 7$, pues $47 \times 7 = 329$ y 48×8 ya se pasa.

6. Situamos c_1 como segunda cifra del resultado, restamos a 344 el número $47 \times 7 = 329$, bajamos las dos cifras siguientes del radicando y, para calcular la cifra siguiente, multiplicamos por 2 el resultado provisional ($27 \times 2 = 54$) y ponemos el resultado en una nueva línea:

$$\begin{array}{r|l}
\sqrt{\;7\,44\,69\,68} & 27 \\
\cline{2-2}
\;\;4 & 47 \\
\cline{1-1}
\;\;34\,4 & 54 \\
\cline{1-1}
\;\;32\,9 & \\
\cline{1-1}
\;\;\;1\,56\,9 &
\end{array}$$

A partir de aquí se repite el mismo proceso. Ahora hay que buscar la mayor cifra c_2 que cumple $54c_2 \times c_2 \leq 1\,569$, pero ahora vamos a explicar usando este paso como ejemplo por qué el algoritmo es correcto. Admitamos que sabemos que

$$N = 2\,700^2 + r,$$

donde $r = 156\,968$ y vamos a ver cómo obtener una cifra más c que refine la aproximación 2 700. Queremos la mayor cifra que cumpla

$$N = (2\,700 + 10c)^2 + r' = 2\,700^2 + 2 \cdot 2\,700 \cdot 10c + 100c^2 + r'.$$

Esto equivale a que

$$r = c(2 \cdot 270 + c) \cdot 100 + r'.$$

Equivalentemente, queremos que $100 \times 54c \times c \leq 156\,968$, lo cual equivale a que

$$54c \times c \leq 1\,569.69$$

y, como son números naturales, a que $54c \times c \leq 1\,569$, que es justo el criterio que estamos siguiendo para determinar la cifra c. En este caso es $c = 2$, y así $542 \times 2 = 1\,084 \leq 1\,569$. Finalmente, el nuevo resto es

$$r' = r - c(2 \cdot 270 + c) \cdot 100 = 156\,968 - 156\,900 = 48\,568,$$

que es justo el valor que calcula el algoritmo descrito.

Ahora observamos que de

$$7\,446\,968 = 2\,728^2 + 4\,984 < 2\,729^2$$

se deduce que

$$74\,469.68 = 272.8^2 + 49.84 < 272.9^2,$$

por lo que

$$\sqrt{74\,469.68} = 272.8\ldots,$$

lo que significa que el algoritmo que acabamos de ver se puede aplicar sin cambio alguno con números decimales, sin más que tener la precaución de formar grupos de dos cifras en el radicando de modo que el punto decimal no quede en medio de un grupo. A su vez, esto significa que si queremos obtener más cifras decimales de una raíz cuadrada, como

$$\sqrt{7\,446\,968} = 2\,728.913337\ldots$$

sólo tenemos que añadir dos ceros a la derecha al radicando por cada nueva cifra que queramos calcular. Por ejemplo:

$$\begin{array}{r|l}
\sqrt{\;2.00\,00\,00} & 1.414 \\
\cline{2-2}
\;\;1 & 24 \\
\cline{2-2}
\;\;100 & 281 \\
\;\;\;96 & 2824 \\
\cline{2-2}
\;\;\;400 & \\
\;\;\;281 & \\
\cline{2-2}
\;\;11\,900 & \\
\;\;11\,296 & \\
\cline{2-2}
\;\;\;3\,836 &
\end{array}$$

Medias Los pitagóricos consideraron tres formas de calcular la media de dos números x e y:

1. La *media aritmética* es $\mathrm{MA} = \dfrac{x+y}{2}$.
2. La *media geométrica* es $\mathrm{MG} = \sqrt{xy}$.
3. La *media harmónica* es $\mathrm{MH} = \dfrac{2}{\frac{1}{x}+\frac{1}{y}} = \dfrac{2xy}{x+y}$.

Vamos a probar que si $x, y > 0$, se cumplen las desigualdades

$$\text{mín}\{x,y\} \le \mathrm{MH} \le \mathrm{MG} \le \mathrm{MA} \le \text{máx}\{x,y\}.$$

En particular, esto muestra que las tres medias son valores intermedios entre x e y, de modo que si $x = y$, las tres medias son iguales al valor común. No perdemos generalidad si suponemos que $x \le y$. Ciertamente, entonces,

$$\mathrm{MA} = \frac{x+y}{2} \le \frac{y+y}{2} = y.$$

La desigualdad

$$\mathrm{MG} = \sqrt{xy} \le \frac{x+y}{2} = \mathrm{MA}$$

es equivalente a $4xy \le (x+y)^2 = x^2 + y^2 + 2xy$, o también a

$$x^2 + y^2 - 2xy = (x-y)^2 \ge 0,$$

lo cual es cierto. Ahora aplicamos la desigualdad que acabamos de probar a los números $1/x$ y $1/y$, de modo que

$$\frac{1}{\sqrt{xy}} = \sqrt{\frac{1}{xy}} \le \frac{1/x + 1/y}{2}.$$

Tomando inversos obtenemos la desigualdad $\mathrm{MH} \le \mathrm{MG}$.

Finalmente, $x \leq y$ implica que $1/y \leq 1/x$, luego $1/x + 1/y \leq 2/x$, luego

$$\mathrm{MH} = \frac{2}{\frac{1}{x} + \frac{1}{y}} \geq \frac{2}{2/x} = x.$$

■

Más aún, todas las desigualdades que hemos probado son estrictas salvo si $x = y$. En efecto, el razonamiento con el que hemos probado $\mathrm{MG} \leq \mathrm{MA}$ muestra que se da la igualdad si y sólo si $(x - y)^2 = 0$, lo que equivale a $x = y$. A su vez, esto implica que $\mathrm{MH} = \mathrm{MG}$ si y sólo si $1/x = 1/y$, que a su vez equivale a $x = y$. Es fácil ver que $\mathrm{MA} = y$ o $\mathrm{MH} = x$ también implican que $x = y$.

1.6 El principio de inducción

En esta revisión de las propiedades básicas del sistema numérico no puede faltar la mención de una potente técnica de demostración como es el llamado principio de inducción matemática, que es una consecuencia inmediata del principio de buena ordenación:

> *Si podemos probar que* 0 *tiene una propiedad y, suponiendo que la cumpla un número natural* n*, podemos probar que también la cumple* $n + 1$*, entonces podemos asegurar que todo número natural tiene la propiedad considerada.*

En efecto, bajo las hipótesis dadas, si existiera un número natural que no cumple la propiedad considerada, podríamos considerar el mínimo m de todos ellos, pero no puede ser $m = 0$, ya que hemos demostrado que 0 sí que cumple la propiedad, luego tiene que ser $m = n + 1$, para cierto $n < m$, pero entonces, por la minimalidad de m, tenemos que n debe cumplir la propiedad, y entonces sabemos demostrar que $m = n + 1$ también la cumple, con lo que tenemos una contradicción.

A veces es más conveniente considerar esta ligera variante:

> *Si, suponiendo que todos los números naturales menores que uno dado* n *cumplen una propiedad, podemos razonar que* n *también la cumple, entonces podemos afirmar que todos los números naturales cumplen dicha propiedad.*

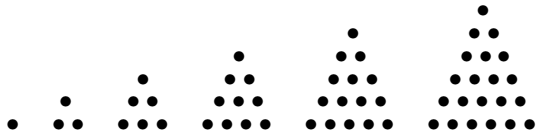
En efecto, si existe un número natural que no cumple la propiedad considerada, por el principio de buena ordenación podemos tomar el mínimo n que no la cumple. Tenemos así un número natural con la propiedad de que todos los precedentes (si los hay) cumplen la propiedad, pero él mismo no la cumple. Así pues, si hemos descartado que este caso pueda darse, también hemos descartado que pueda haber un número natural que no cumpla la propiedad.

Cuando, en un razonamiento por inducción, suponemos que n cumple la propiedad correspondiente, o que los números menores que n la cumplen, dicha hipótesis recibe el nombre de *hipótesis de inducción.*

Sumas de potencias Veamos algunos ejemplos de uso del principio de inducción. Los números de la forma

$$1 + 2 + 3 + \cdots + n$$

se llaman *números triangulares,* por razones obvias:



La conocida fórmula

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

puede probarse fácilmente por inducción, aunque hay un argumento más directo,[18] por ejemplo, para $n = 6$ tenemos que

$$1 + 2 + 3 + 4 + 5 + 6 = (1 + 6) + (2 + 5) + (3 + 6) = 7 + 7 + 7 = 3 \cdot 7.$$

En general, si $n = 2k$ es par, al agrupar los sumandos de dos en dos, obtenemos k sumandos iguales a $n + 1$, luego la suma es $k(n + 1) = n(n + 1)/2$.

Por otra parte, para $n = 7$ podemos razonar así:

$$1 + 2 + 3 + 4 + 5 + 6 + 7 = (0 + 7) + (1 + 6) + (2 + 5) + (3 + 4)$$

$$= 7 + 7 + 7 + 7 = 7 \cdot 4.$$

Y en general, si $n = 2k+1$, añadiendo un sumando 0 obtenemos $2k+2$ sumandos que, agrupados en pares, dan $k + 1$ sumados iguales a n, luego la suma es $n(k + 1) = n(n + 1)/2$.

Ejercicio: Probar la fórmula $1 + 3 + 5 + \cdots + 2k - 1 = k^2$.

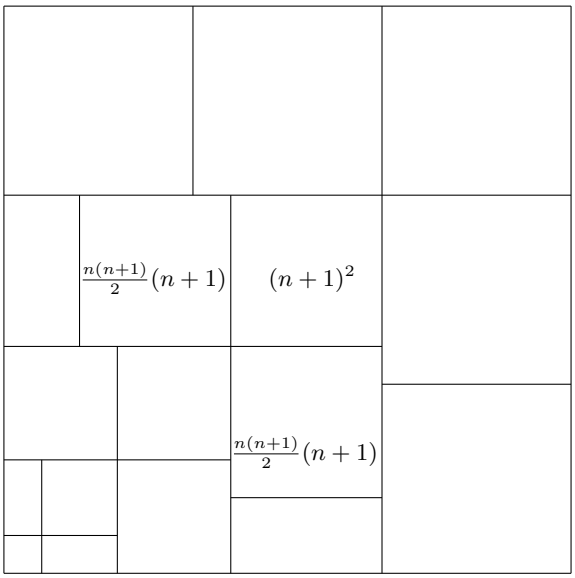
Ejercicio: Probar que la suma de dos números triangulares consecutivos es un cuadrado.

Más interesante es la fórmula conocida como *teorema de Nicómaco:*

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4} = (1 + 2 + 3 + \cdots + n)^2.$$

La figura siguiente muestra la geometría subyacente a la prueba (para $n = 3$):

[18] En realidad, el argumento que damos emplea pasos (como las reordenaciones de sumandos) que para ser completamente formalizados requieren argumentos inductivos, pero podemos distinguir entre argumentos inductivos que son intuitivamente evidentes y pueden entenderse sin necesidad de formular explícitamente una inducción (como las reordenaciones de sumandos) y argumentos inductivos que requieren explicitar la inducción.



Razonamos por inducción tomando como hipótesis de inducción que se cumple para todos los números menores que m. Ahora suponemos que $m \geq 1$, con lo que $m = n+1$ y tenemos que probar la fórmula para m. Si $n = 0$ ésta se reduce a $1^3 = 1^2$, que es obviamente cierta. En caso contrario, usamos la relación

$$(A+B)^2 = A^2 + B^2 + 2AB,$$

con

$$A = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}, \qquad B = n + 1.$$

La hipótesis de inducción nos da que

$$(1 + 2 + 3 + \cdots + n + 1)^2 = 1^3 + 2^3 + 3^3 + \cdots + n^3 + (n+1)^2 + n(n+1)^2$$

$$= 1^3 + 2^3 + 3^3 + \cdots + (n+1)^3,$$

y esto prueba que la fórmula es válida en general.

En la figura vemos que la diferencia entre el cuadrado de lado $1+\cdots+(n+1)$ y el cuadrado de lado $1 + \cdots + n$ consta de uncuadrado de lado $n + 1$ y dos rectángulos formados, o bien por $n/2$ cuadrados de lado $n+1$ (cuando n es par) o bien por $(n-1)/2$ cuadrados y medio de lado $n+1$, luego en total la diferencia de las áreas equivale a la de $n + 1$ cuadrados de lado $n + 1$.

Ejercicio: Probar por inducción que

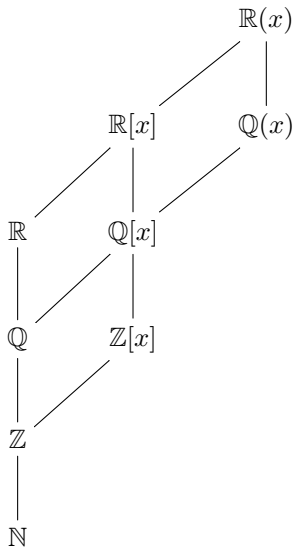
$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} = \frac{2n^3 + 3n^2 + n}{6}.$$

1.7 Notación conjuntista

El lector familiarizado con la notación conjuntista habitual en matemáticas habrá notado que en las páginas precedentes la hemos evitado en gran medida. No obstante, en ocasiones resultará útil hacer uso de ella. Para empezar, los conjuntos numéricos que hemos estudiado tienen nombres tradicionales:

- $\mathbb{N}$ El conjunto de los números naturales
- $\mathbb{Z}$ El conjunto de los números enteros (un dominio íntegro ordenado)
- $\mathbb{Q}$ El conjunto de los números racionales (cuerpo ordenado)
- $\mathbb{R}$ El conjunto de los números reales (cuerpo ordenado)

Todavía podemos extenderlos más considerando sus anillos de polinomios y sus cuerpos de fracciones algebraicas. Notemos la igualdad entre los cuerpos de fracciones algebraicas $\mathbb{Z}(x) = \mathbb{Q}(x)$, pues al considerar las fracciones de todos los polinomios, en particular estamos considerando las fracciones de números enteros, por lo que $\mathbb{Z}(x)$ contiene a $\mathbb{Q}$ y a todos los polinomios y cocientes de polinomios con coeficientes racionales:



La notación $a \in A$, $a \notin A$ expresa que a es (o no es) uno de los elementos del conjunto A. Por ejemplo, $-5 \in \mathbb{Z}$, $-5 \notin \mathbb{N}$. También será útil considerar en ocasiones productos cartesianos $A \times B$ de dos o más conjuntos, formados por todos los pares (o ternas, o cuádruplas, etc.) de la forma (a, b), con $a \in A$ y $b \in B$, así como aplicaciones (o funciones) $f : A \longrightarrow B$ que a cada elemento $a \in A$ le hacen corresponder un elemento $f(a) \in B$.

Por ejemplo, la suma de números enteros puede verse como una aplicación $+ : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z}$, pero en general evitaremos este tipo de notación salvo que realmente sea clarificadora.

Capítulo II

La aritmética de la escuela

La frontera entre el álgebra y la aritmética es difusa, pero, a grandes rasgos y superficialmente, podríamos decir que todo lo que tiene que ver con sumar, restar, multiplicar y dividir es álgebra, mientras que lo relacionado con múltiplos, divisores, primos, etc. es aritmética. En el capítulo anterior repasamos el álgebra más básica y ahora vamos a repasar la aritmética básica. Por ejemplo, un resultado que sin duda el lector aprendería en la escuela es que todo número natural (mayor que 1) se descompone de forma única el producto de números primos. Esto es lo que se conoce como el teorema fundamental de la aritmética y será uno de los resultados que demostraremos aquí. Más aún, veremos que tiene un análogo en anillos de polinomios. Igual que podemos descomponer en factores primos $60 = 2^2 \cdot 3 \cdot 5$, resulta que

$$x^4 - 1 = (x-1)(x+1)(x^2+1)$$

es la descomposición en factores primos del polinomio $x^4 - 1$ en el anillo $\mathbb{Q}[x]$. La aritmética en un anillo como $\mathbb{Q}[x]$ tiene algunas características distintas de la aritmética en $\mathbb{N}$, pero veremos que, si adoptamos la perspectiva algebraica adecuada, ambas tienen esencialmente las mismas características.

2.1 La aritmética de $\mathbb{N}$

Recordemos que un número natural m divide a otro n (en signos $m \mid n$) si existe un tercero r tal que $n = mr$.

Hay dos casos triviales de divisibilidad, y es que 0 es divisible entre todos los números naturales (pues $0 = n \cdot 0$) y 1 divide a todos los números naturales (pues $n = 1 \cdot n$).

Máximo común divisor Excluyendo el caso $n = 0$, es claro que si $m \mid n$ entonces $m \leq n$, por lo que cada número natural no nulo n tiene un número finito de divisores, el mayor de los cuales es el propio n.

Por lo tanto, dados dos números naturales no nulos m y n, podemos considerar lo que se llama el *máximo común divisor* de m y n, que representaremos por (m, n), y que no es sino lo que su propio nombre indica: el mayor de los números que dividen tanto a m como a n.

Notemos que $(m, n) = (n, m)$. Cuando $(m, n) = 1$, es decir, cuando m y n no tienen más divisores comunes que el 1, se dice que son *primos entre sí.*

Existe un algoritmo muy eficiente para calcular el máximo común divisor de dos números naturales que aparece ya en los *Elementos,* de Euclides, y que se basa en el hecho siguiente:

$$(a, b) = (a, b \pm ac).$$

(En principio podemos suponer que $ac \leq b$ si tomamos el signo negativo, para que $b - ac$ sea un número natural, pero en realidad la definición de divisibilidad vale trivialmente para números enteros y en ningún momento vamos a usar esta hipótesis.)

En efecto, basta ver que los divisores comunes de a y b son los mismos que los divisores comunes de a y $b \pm ac$.

Si $d \mid a$ y $d \mid b$, entonces $a = a'd$, $b = b'd$, luego

$$b \pm ac = b'd \pm a'dc = (b' \pm a'c)d,$$

luego $d \mid b \pm ac$ (y, por supuesto, $d \mid a$). Recíprocamente, si $d \mid a$ y $d \mid b \pm ac$, podemos concluir que $d \mid a$ y $d \mid b$, bien por un razonamiento análogo o bien porque es lo mismo que ya hemos demostrado, ya que $b = (b \pm ac) \mp ac$.

En particular, si $a \leq b$ y dividimos $b = ac + r$, con $0 \leq r < a$, tenemos que

$$(a, b) = (a, r).$$

Esto significa que, para calcular el máximo común divisor de dos números, podemos cambiar el mayor de ambos por el resto de su división entre el menor.

Ejemplo Vamos a calcular $(4\,070, 3\,626)$ mediante al algoritmo de Euclides.

$$\begin{array}{r|r} & 4\,070 \\ 1 & 3\,626 \\ 8 & 444 \\ 6 & 74 \\ & 0 \end{array}$$

Para ello situamos los dos números uno debajo del otro, con el mayor arriba. Realizamos la división euclídea,

$$4\,070 = 3\,626 \cdot 1 + 444$$

y escribimos el resto debajo de los dos números. El razonamiento precedente muestra que

$$(4\,070, 3\,626) = (3\,626, 444).$$

Ahora repetimos el proceso: dividimos $3\,626 = 444 \cdot 8 + 74$, con lo que

$$(4\,070, 3\,626) = (3\,626, 444) = (444, 74),$$

y una última división nos lleva a que

$$(4\,070, 3\,626) = (3\,626, 444) = (444, 74) = (74, 0) = 74.$$

Notemos que, en general, los restos que vamos escribiendo forman una sucesión estrictamente decreciente, luego tras un número finito de pasos tenemos que llegar necesariamente a 0 y entonces tenemos calculado el máximo común divisor. ∎

En realidad el algoritmo de Euclides es interesante porque, refinándolo ligeramente, nos da una información muy importante:

Relación de Bezout *Si a y b son números naturales no nulos, existen números enteros u y v tales que $(a, b) = ua + vb$.*

En efecto, pongamos que tenemos dos números naturales $x = u_1a + v_1b$, $y = u_2a + v_2b$ y que dividimos $x = yc + r$, con $0 \leq r < y$. Entonces

$$r = x - yc = u_1a + v_1b - (u_2a + v_2b)c = (u_1 - u_2c)a + (v_1 - v_2c)b.$$

Así, dados dos números naturales, como $a = 4\,070$, $b = 3\,626$, podemos expresarlos trivialmente como

$$4\,070 = 1 \cdot a + 0 \cdot b, \qquad 3\,626 = 0 \cdot a + 1 \cdot b,$$

y en la tabla con la que aplicamos el algoritmo de Euclides podemos anotar estos valores triviales de u y v:

c	r	u	v
	4 070	1	0
1	3 626	0	1
8	444	1	−1
6	74	−8	9
	0		

Ahora, el cálculo precedente muestra que si, una vez determinado el cociente $c = 1$ que hemos anotado en la segunda fila a la izquierda, escribimos en la tercera fila los números que resultan de restarle a los de la primera los correspondientes en la segunda multiplicados por $c = 1$, los valores $u = 1$, $v = -1$ cumplen que

$$444 = 1 \cdot a + (-1) \cdot b.$$

Igualmente, una vez calculado el nuevo cociente $c = 8$, escribimos en la cuarta fila los números que resultan de restar a los de la segunda fila los de la tercera multiplicados por $c = 8$, y así obtenemos los números u y v que cumplen

$$74 = -8a + 9b = -8 \cdot 4\,070 + 9 \cdot 3\,626.$$

Es obvio que el procedimiento es general, es decir, que este algoritmo siempre nos proporciona enteros u y v que cumplen la relación de Bezout. ∎

Como primera aplicación de la relación de Bezout observamos que si a y b son números naturales y d es un divisor común, la definición de máximo común divisor sólo nos asegura que $d \leq (a, b)$, pero la relación de Bezout nos asegura que, de hecho, $d \mid (a, b)$.

Así, pues, ahora sabemos que (a, b) no sólo es mayor o igual que todos los divisores comunes de a y b, sino que de hecho es múltiplo de todos los divisores comunes de a y b.

Pero la consecuencia fundamental de la relación de Bezout tiene que ver con los números primos.

Números primos Observemos que todo número natural $n \geq 2$ tiene al menos dos divisores, a saber, 1 y n.

> Se dice que un número natural $p \geq 2$ es *primo* si sus únicos divisores son 1 y p. Los números $n \geq 2$ que no son primos se llaman *compuestos.*

Es claro que el menor número primo es $p = 2$. Como 3 no es divisible entre 2, resulta que también es primo. En general, una forma sencilla de encontrar números primos es aplicar la llamada *criba de Eratóstenes.* Consiste en escribir todos los números hasta uno dado, por ejemplo, hasta 100, tachar el 0 y el 1, que por definición no son primos, luego tachamos todos los múltiplos de 2 (menos el propio 2) y así el menor número que quede sin tachar (el 3) será primo. Luego tachamos todos los múltiplos de 3 (menos el 3) y el primer número que quede sin tachar (el 5) será primo, y así sucesivamente. Los números que sobreviven al proceso de criba son todos los primos en el intervalo considerado. El resultado es que hay exactamente 25 primos menores que 100:

	2	3		5		7			
11		13				17		19	
		23						29	
31						37			
41		43				47			
		53						59	
61						67			
71		73						79	
		83						89	
						97			

En realidad, una vez eliminados todos los múltiplos de 7, ya podíamos asegurar que no quedaba ningún número compuesto menor que 100. Enseguida entenderemos por qué:

1. *Si $n \geq 2$, entonces el menor divisor p de n tal que $p \geq 2$ es primo.*

   En efecto, si $d \mid p$, entonces $d \mid n$ y $d \leq p \leq n$, luego, o bien $d = 1$, o bien, si $d \geq 2$, por la minimalidad de p, tiene que ser $d = p$, y esto significa que p es primo.

2. *Si $n \geq 2$ no es primo, entonces tiene un divisor primo p tal que $p \leq \sqrt{n}$.*

   Basta considerar el menor divisor primo p de n. Si n no es primo, entonces $p \neq n$, luego $n/p \geq 2$, luego n/p tiene un divisor primo q, luego $pq \mid n$, luego $p \leq q$ por la minimalidad de p, luego $p^2 \leq pq \leq n$.

   Así, todo número compuesto $n \leq 100$ es divisible entre un primo p tal que $p^2 \leq n \leq 100 = 10^2$, luego $p \leq 10$, luego $p = 2, 3, 5, 7$. Por lo tanto, al aplicar la criba de Eratóstenes hasta 100, todos los números compuestos quedan eliminados una vez tachamos los múltiplos de 7, como afirmábamos.

3. *Existen infinitos primos.*

   Dado cualquier número natural $n \geq 2$, el número $n! + 1$ debe tener un divisor primo p, pero no puede ser $p \leq n$, pues entonces $p \mid n!$ y $p \mid n! + 1$, de donde se sigue inmediatamente que $p \mid 1$, lo cual es absurdo. Así pues, $p > n$, y hemos probado que existen primos mayores que cualquier número natural n prefijado.

4. *Un número natural $p \geq 2$ es primo si y sólo si cuando $p \mid ab$, entonces $p \mid a$ o $p \mid b$.*

   En efecto, si p es primo y $p \mid ab$, entonces $(p, a) = p$ o bien $(p, a) = 1$, pues p no tiene más divisores que 1 y p. En el primer caso tenemos que $p \mid a$, mientras que en el segundo podemos aplicar la relación de Bezout, que nos da números enteros u y v tales que $1 = up + va$, luego $b = upb + vab$, y como $p \mid ab$, de aquí se sigue que $p \mid b$.

   Recíprocamente, si p cumple esta propiedad y $a \mid p$, entonces existe un b tal que $p = ab$, luego $p \mid a$ o $p \mid b$. Pero, como $a \mid p$ y $b \mid p$, esto implica que $p = a$ o bien $p = b$, y en el segundo caso $a = 1$, luego p no tiene más divisores que 1 y p.

Observemos que esta última propiedad admite una generalización obvia:

Si p es primo y $p \mid a_1 \cdots a_n$, entonces existe un i tal que $p \mid a_i$.

En efecto, tomemos, por simplicidad $n = 4$. Si $p \mid abcd$, entonces, por la propiedad anterior, $p \mid a$ o bien $p \mid bcd$, pero en el segundo caso $p \mid b$ o bien $p \mid cd$, y en el último caso $p \mid c$ o $p \mid d$. Así pues, tiene que darse uno de los cuatro casos $p \mid a$ o $p \mid b$ o $p \mid c$ o $p \mid d$. Es claro que el argumento vale para cualquier número de factores.

En la demostración de esta propiedad ha sido fundamental la relación de Bezout, y esta propiedad es a su vez esencial para demostrar la unicidad de la descomposición en primos:

Teorema fundamental de la aritmética *Todo número natural $n \geq 2$ se descompone de forma única (salvo el orden de los factores) en producto de números primos.*

La unicidad salvo el orden quiere decir que podemos tener descomposiciones como

$$20 = 2 \cdot 2 \cdot 5 = 2 \cdot 5 \cdot 2,$$

que no son literalmente la misma. Ahora bien, si agrupamos los factores primos en potencias, así:

$$20 = 2^2 \cdot 5$$

y ordenamos los primos de menor a mayor, entonces la descomposición es única.

DEMOSTRACIÓN: La prueba de que todo número $n \geq 2$ se puede descomponer en factores primos es inmediata. Sabemos que n tiene un factor primo p_1, de modo que $n = p_1 n_1$. Si $n_1 \geq 2$, entonces tiene un factor primo p_2, de modo que $n = p_1 p_2 n_2$, si $n_2 \geq 2$, entonces tiene un factor primo p_3, de modo que $n = p_1 p_2 p_3 n_3$, y así vamos obteniendo una sucesión estrictamente decreciente

$$n > n_1 > n_2 > n_3 > \cdots$$

que se podrá prolongar mientras $n_i \geq 2$, luego tras un número finito de pasos tiene que ser $n_r = 1$, y entonces llegamos a que $n = p_1 \cdots p_r$ queda descompuesto en producto de factores primos.

La parte delicada es justificar que la descomposición es única, es decir, que no puede suceder que

$$3 \cdot 7^2 \cdot 23 = 2 \cdot 3^5 \cdot 7$$

En este caso $3\,402 \neq 3\,381$, pero, ¿no podría darse una coincidencia de este tipo con otras descomposiciones en factores primos? La respuesta es que no. En primer lugar, no puede ser que un mismo primo aparezca en una descomposición y no en otra. En efecto, si tenemos dos descomposiciones

$$p_1 \cdots p_r = q_1 \cdots q_s$$

en factores primos, entonces cada p_i divide a $q_1 \cdots q_s$, luego, según hemos visto (y en este paso es crucial la relación de Bezout) tiene que habar un j tal que $p_i \mid q_j$, pero como q_j es primo y $p_i \neq 1$, tiene que ser $p_i = q_j$. En otras palabras, todos los primos que aparecen a la izquierda tienen que aparecer a la derecha, y viceversa.

Pero aún tenemos que descartar la posibilidad de que hubiera dos descomposiciones en factores primos de un mismo número de la forma

$$3^5 \cdot 11^2 \cdot 19 = 3^3 \cdot 11^3 \cdot 19,$$

con los mismos primos, pero repetidos un número diferente de veces. Ahora bien, esto tampoco es posible, porque en tal caso, tomamos un primo p que aparezca m veces en una descomposición y $n > m$ veces en otra, y dividimos ambas descomposiciones entre p^m. En nuestro ejemplo, si se diera la igualdad que hemos planteado, dividiendo entre 3^3 obtendríamos

$$3^2 \cdot 11^2 \cdot 19 = 11^3 \cdot 19,$$

y así tendríamos dos descomposiciones en factores primos de un mismo número con un primo p que aparece en una y no en la otra, cuando ya hemos visto que eso es imposible.

Esto obliga a que en dos descomposiciones en factores primos de un mismo número, tengan que aparecer los mismos primos repetidos el mismo número de veces, y esto es lo que afirma el teorema fundamental. ∎

La aritmética de los números naturales está determinada por las descomposiciones en factores primos. Por ejemplo, si conocemos las descomposiciones de dos números, podemos saber inmediatamente si uno divide o no al otro. Consideremos, por ejemplo,

$$m = 2 \cdot 3^3 \cdot 11^2, \qquad n = 2^3 \cdot 3^3 \cdot 5^2 \cdot 7 \cdot 11^4, \qquad r = 2^2 \cdot 3^2 \cdot 11^3.$$

Es inmediato que $m \mid n$, pero $m \nmid r$. Para ello basta observar que todos cada primo aparece en la descomposición de m con exponente menor o igual que en n, por lo que podemos obtener n a partir de m multiplicándolo por los primos que faltan:

$$n = m \cdot 2^2 \cdot 3 \cdot 5^2 \cdot 7 \cdot 11^2.$$

Por el contrario, el exponente de 3 en m es mayor que el de r, y esto implica que $m \nmid r$. Aquí es esencial la unicidad de las descomposiciones en primos, pues el argumento es que, si m dividiera a r, tendríamos que $r = mc$, para cierto c, y entonces la descomposición en factores primos de r sería (por la unicidad) el producto de la descomposición en factores primos de m por la de c y, como en la de m aparecen tres treses, en la de r tendría que haber al menos tres treses, y no los hay. En general:

> *Si m, $n \geq 2$ son números naturales, se cumple que $m \mid n$ si y sólo si el exponente de cada primo en la descomposición en factores primos de m es menor o igual que su exponente en la descomposición en factores primos de n.*

Esto nos permite a su vez calcular el máximo común divisor y el mínimo común múltiplo de unos números naturales dados. Hemos definido el máximo común divisor para el caso de dos números, pero la definición se generaliza de forma obvia:

> Si $m_1, \ldots, m_n$ son números naturales no nulos, su *máximo común divisor* es el mayor de los divisores comunes a todos ellos, y su *mínimo común múltiplo* es el mínimo de los múltiplos comunes.

Notemos que siempre existe un divisor común (1) y un múltiplo común $m_1 \cdots m_n$, y que los divisores comunes están acotados por cualquiera de los números dados, por lo que siempre existe el máximo común divisor y el mínimo común múltiplo. La forma de calcularlos a partir de las descomposiciones en factores primos es la siguiente:

El máximo común divisor de unos números naturales $m_1, \ldots, m_n$ no nulos es el producto de los divisores primos comunes elevados al menor exponente con que aparecen en ellos. El mínimo común múltiplo es el producto de todos sus divisores primos (comunes y no comunes) elevados al mayor exponente con que aparecen en ellos.

Por ejemplo, dados los números

$$m_1 = 2 \cdot 3^4 \cdot 7^3, \quad m_2 = 3^2 \cdot 7 \cdot 13^3, \quad m_3 = 2^2 \cdot 3^3 \cdot 7^2 \cdot 23,$$

su máximo común divisor d y su mínimo común múltiplo m son

$$d = 3^2 \cdot 7, \qquad m = 2^2 \cdot 3^3 \cdot 7^3 \cdot 13^3 \cdot 23.$$

En efecto, se cumple que $d \mid m_1$, $d \mid m_2$ y $d \mid m_3$, porque todos los primos dividen a d con exponente menor o igual que a cada uno de los m_i y, si c es otro divisor común de los m_i, los exponentes de 3 y 7 en c tienen que ser menores o iguales que sus exponentes en cada m_i, luego en particular tienen que ser menores o iguales que 3 y 1, respectivamente, que son los mínimos valores de dichos exponentes, luego, no sólo podemos afirmar que $c \leq d$, sino que, de hecho, se cumple que $c \mid d$.

Igualmente es claro que $m_1 \mid m$, $m_2 \mid m$ y $m_3 \mid m$, porque todos los primos dividen a los m_i con exponente menor o igual que a m, y si c es otro múltiplo común a todos los m_i, entonces el exponente de cada uno de los cinco primos que dividen a m tiene que ser mayor o igual en c que en cada m_i, luego tiene que ser mayor o igual que su exponente en m, que es el máximo de estos, luego no sólo $m \leq c$, sino que de hecho $m \mid c$.

Es claro que estos razonamientos se aplican a cualesquiera que sean los números dados, no necesariamente los de este ejemplo, con la única salvedad de que si $m_i = 1$ para algún i, entonces directamente podemos concluir que el máximo común divisor es 1, ya que $m_i = 1$ no tiene más divisores, y el mínimo común múltiplo es el mismo que el que resulta de eliminar dicho m_i. Y si todos los m_i fueran iguales a 1 entonces el mínimo común múltiplo sería 1.

Con esto hemos probado que el máximo común divisor y el mínimo común múltiplo cumplen algo más fuerte de lo que requiere la definición:

El máximo común divisor de unos números dados $m_1, \ldots, m_n$ es múltiplo de todos sus divisores comunes. El mínimo común múltiplo divide a todos los múltiplos comunes.

Ejercicio: Probar que si d y m son el máximo común divisor y el mínimo común múltiplo de dos números naturales a y b, entonces $ab = dm$.

Veamos otro ejemplo de relación entre un concepto aritmético y la descomposición en factores primos:

Los números naturales de la forma n^2 (donde n es un número natural) se llaman *cuadrados perfectos,* los de la forma n^3 se llaman *cubos perfectos* y, en general, los de la forma n^m se llaman potencias m-simas perfectas.

Es obvio que un número natural r es una potencia m-sima perfecta si y sólo si todos los exponentes de los primos que lo dividen son múltiplos de m, y en tal caso, el número n que cumple $r = n^m$ es el que resulta de dividir dichos exponentes entre m.

Por ejemplo, $324 = 2^2 \cdot 3^4$ es un cuadrado perfecto, ya que es de la forma $324 = (2 \cdot 3^2)^2 = 18^2$. En cambio, $108 = 2^2 \cdot 3^3$ no es un cuadrado perfecto, ya que el exponente de 3 en su descomposición en factores es impar.

Esto tiene una consecuencia algebraica. Aunque no hemos hablado de números reales, el lector sabrá sin duda que $\mathbb{R}$ es un cuerpo que contiene a $\mathbb{Q}$ en el que existen números como $\sqrt{2}$ o $\sqrt[3]{7}$, y es conocido que son números irracionales. He aquí la prueba:

> *Si n es un número natural que no es una potencia m-sima perfecta, entonces no existe ningún número racional r tal que $r^m = n$.*

En efecto, si existiera tal r, cambiándole el signo si fuera preciso podríamos tomarlo positivo, luego podríamos expresarlo como $r = a/b$, donde a y b son números naturales no nulos. Tendríamos entonces que $(a/b)^m = n$ o, equivalentemente, que $a^m = nb^m$.

Como n no es una potencia m-sima perfecta, tiene un divisor primo, digamos p, cuyo exponente en n no es múltiplo de m. Pongamos que $n = p^u x$, $a = p^v y$, $b = p^w z$, donde $p \nmid x$, $p \nmid y$, $p \nmid z$ y $m \nmid u$. Entonces tenemos que $p^{mv} y^m = p^u x p^{mw} z^m$, luego por la unicidad de la factorización tiene que ser $mv = u + mw$, de donde $m \mid u$, en contra de lo supuesto. ∎

2.2 Aplicaciones de la factorización única en $\mathbb{Z}$

Ecuaciones diofánticas lineales Las ecuaciones diofánticas más sencillas (con al menos dos variables, pues en caso contrario son triviales) son las de la forma

$$ax + by = c,$$

donde a, b, c son números enteros tales que $ab \neq 0$. Es obvio que una condición necesaria para que una ecuación de este tipo tenga solución es necesario que $d = (a, b) \mid c$, en cuyo caso las soluciones son las mismas que las de

$$(a/d)x + (b/d)y = c/d,$$

por lo que no perdemos generalidad si suponemos que $(a, b) = 1$. En tal caso, el teorema de Bezout nos da enteros u y v tales que $au + bv = 1$, por lo que $(x_0, y_0) = (uc, vc)$ es una solución de la ecuación.

Si (x, y) es una solución arbitraria, entonces $a(x - x_0) + b(y - y_0) = 0$, luego $a(x - x_0) = -b(y - y_0)$. Como a y b son primos entre sí, esto implica que $a \mid (y - y_0)$ y $b \mid (x - x_0)$. Pongamos que $x - x_0 = kb$, $y - y_0 = ra$. Entonces $kab + rba = 0$, luego $k + r = 0$. En suma:

$$(x, y) = (x_0 + kb, y_0 - ka),$$

y es claro que todo par de enteros (x, y) de esta forma, para un k arbitrario, es solución de la ecuación. Así pues, hemos demostrado lo siguiente:

Teorema 2.1 *Una ecuación de la forma $ax+by = c$ (donde a, b, c son números enteros y $ab \neq 0$) tiene soluciones enteras si y sólo si $d = (a, b) \mid c$, y en tal caso, las soluciones son todos los pares*

$$(x, y) = (x_0 + kb/d, y_0 - ka/d),$$

donde (x_0, y_0) es una solución particular arbitraria de la ecuación.

Ejemplo *Cinco marineros llegan a una isla desierta y durante el día recogen todos los cocos que pueden encontrar, para asegurarse de que no les faltará agua. Luego acuerdan que al día siguiente los repartirán a partes iguales. Sin embargo, durante la noche, uno de los marineros, temiendo que el reparto no vaya a ser equitativo, decide llevarse su parte en secreto: divide los cocos en cinco montones iguales y se encuentra con que le sobra uno, arroja a los monos el que le sobra y se lleva la quinta parte. Poco después, otro de los marineros tiene la misma idea y de nuevo reparte los cocos en cinco partes iguales, se encuentra con que le sobra uno, lo arroja a los monos y se lleva la quinta parte. Lo mismo hacen, sucesivamente, los tres marineros restantes. Todos ellos se encuentran con que sobra un coco al hacer el reparto y lo arrojan a los monos. A la mañana siguiente, sin que ninguno confiese que ya se ha llevado una parte de las provisiones, los cocos restantes se dividen en cinco partes iguales (esta vez no sobra ninguno) y cada marinero se queda con una de ellas. ¿Cuántos cocos habían recogido los marineros?*

Solución: Llamemos x al número de cocos inicial. El primer marinero se lleva $(x-1)/5$ de ellos y deja $4(x-1)/5$. Igualmente el segundo marinero deja

$$\frac{4}{5}\left(\frac{4x-4}{5} - 1\right) = \frac{16x-36}{25},$$

el tercer marinero deja

$$\frac{4}{5}\left(\frac{16x-36}{25} - 1\right) = \frac{64x-244}{125},$$

el cuarto marinero deja

$$\frac{4}{5}\left(\frac{64x-244}{125} - 1\right) = \frac{256x-1\,476}{625},$$

y el quinto marinero deja

$$\frac{4}{5}\left(\frac{256x-1\,476}{625} - 1\right) = \frac{1\,024x-8\,404}{3\,125} \text{ cocos.}$$

El número restante de cocos es de la forma $5y$, luego tiene que cumplirse la ecuación

$$\frac{1\,024x-8\,404}{3\,125} = 5y$$

o equivalentemente,

$$1\,024x - 15\,625y = 8\,404.$$

Puesto que $1\,024 = 2^{10}$ y $15\,625 = 5^6$, tenemos que los coeficientes son primos entre sí, luego la ecuación tiene soluciones enteras. Para encontrar una de ellas, tenemos que resolver la ecuación

$$1\,024x - 15\,625y = 1,$$

para lo cual aplicamos el algoritmo de Euclides:

c	r	u	v
	15 625	1	0
15	1 024	0	1
3	265	1	−15
1	229	−3	46
6	36	4	−61
2	13	−27	412
1	10	58	−885
3	3	−85	1 297
	1	313	−4 776

Así concluimos que $1\,024(-4776) - 15\,625 \cdot (-313) = 1$, luego, multiplicando por 8 404, resulta que

$$(x_0, y_0) = (-40\,137\,504, -2\,630\,452)$$

es una solución de la ecuación. Las demás serán de la forma

$$(x, y) = (-40\,137\,504 + 15\,625k, -2\,630\,452 + 1\,024k)$$

Para que las componentes sean positivas necesitamos que

$$k \geq \frac{40\,137\,504}{15\,625} = 2\,568.8, \qquad k \geq \frac{2\,630\,452}{1\,024} = 2\,568.8$$

Con $k = 2\,569$ obtenemos la solución

$$(x_1, y_1) = (3\,121, 204),$$

que es la menor solución natural. La siguiente ya sería $(x_2, y_2) = (18\,746,\ 228)$, que supondría que cada marinero tendría que haber recogido una media de 3 749 cocos, así que nos quedamos con $x = 3\,121$ como el número más razonable de cocos que podían haber recogido y $5 \cdot 204 = 1\,020$ como el número de cocos repartido finalmente. ∎

Ejercicio: Un hombre cobra un cheque, pero el cajero que se lo paga, se equivoca y le da tantos euros como céntimos tendría que haberle dado y tantos céntimos como euros (por ejemplo, si el importe del cheque hubiera sido de de 60.45 euros, el cajero le habría pagado 45.60 euros). Después de haberse gastado 5 céntimos, el hombre advierte el error, pues comprueba que tiene justo el doble de dinero que debería haber cobrado. ¿Cuál era el importe del cheque?

Ternas pitagóricas Diofanto trató en su *Aritmética* el problema de encontrar ternas de números naturales no nulos x, y, z tales que $x^2 + y^2 = z^2$. Estas ternas se llaman *ternas pitagóricas,* pues según el teorema de Pitágoras permiten construir triángulos rectángulos con lados enteros. Los egipcios las usaban para construir ángulos rectos en arquitectura. Entre los ejemplos más conocidos están

$$3^2 + 4^2 = 5^2, \qquad 5^2 + 12^2 = 13^2, \qquad 7^2 + 24^2 = 25^2,$$

paro, ¿cómo encontrarlas todas?

En primer lugar notamos que si (x, y, z) es una terna pitagórica, también lo es (mx, my, mz) para cualquier número m y, recíprocamente, dada una terna pitagórica (x, y, z), podemos dividir sus componentes por su máximo común divisor para obtener otra que cumpla además $(x, y, z) = 1$. Una terna cuyos elementos no tengan divisores comunes se llama *primitiva.* Si encontramos un método para hallar todas las ternas primitivas, las restantes se obtienen multiplicándolas por números arbitrarios, luego el problema está resuelto. Las tres ternas pitagóricas que hemos dado de ejemplo son todas primitivas.

Ante todo notemos que un divisor primo de dos de las componentes de una terna pitagórica, divide a la tercera. Por ejemplo, si $p \mid x$ y $p \mid z$, entonces $p \mid z^2 - x^2$, con lo que $p \mid y^2$ y por lo tanto $p \mid y$. Esto significa que, en realidad, las componentes de una terna pitagórica primitiva son primas entre sí dos a dos.

En segundo lugar, observamos que esto obliga a que al menos dos componentes de una terna pitagórica sean impares, y la ecuación que satisfacen implica a su vez que la tercera sea par. Más precisamente, tiene que cumplirse que z sea impar, pues si fuera $z = 2z'$ y, por consiguiente, $x = 2x' + 1$, $y = 2y' + 1$, tendríamos que

$$4z^2 = (2x' + 1)^2 + (2y' + 1)^2 = 4x'^2 + 4x' + 4y'^2 + 4y' + 2,$$

de donde concluimos que $4 \mid 2$, lo cual es absurdo.

Así pues, en una terna pitagórica primitiva (x, y, z) tiene que ser par x o y. No perdemos generalidad si consideramos únicamente ternas pitagóricas primitivas en las que y es par.

Teorema 2.2 *Toda terna pitagórica primitiva (x, y, z) (en la que y es par) es de la forma*

$$(x, y, z) = (p^2 - q^2, 2pq, p^2 + q^2),$$

donde $q < p$ son números naturales primos entre sí de paridad opuesta. Recíprocamente, toda terna que admita esta expresión es una terna pitagórica primitiva.

La tabla 2.2 contiene las ternas correspondientes a los valores de $p \leq 7$. En una tablilla cuneiforme aproximadamente del año 1500 a.C. se ha encontrado una enumeración de ternas pitagóricas, entre las cuales se encontraba $(4\,961, 6\,480, 8\,161)$. Se obtiene con $p = 81$ y $q = 40$.

Tabla 2.1: Ternas pitagóricas

p	q	x	y	z
2	1	3	4	5
3	2	5	12	13
4	1	15	8	17
4	3	7	24	25
5	2	21	20	29
5	4	9	40	41
6	1	35	12	37
6	5	11	60	61
7	2	45	28	53
7	4	33	56	65
7	6	13	84	85

DEMOSTRACIÓN: Consideremos una terna pitagórica primitiva arbitraria (x, y, z) en la que y es par, luego x y z son impares. Entonces $z + x$, $z - x$ son ambos pares. Digamos que $y = 2u$, $z + x = 2v$, $z - x = 2w$. Ahora

$$y^2 = z^2 - x^2 = (z + x)(z - x),$$

luego $u^2 = vw$, $v > 0$, $w > 0$. Por otro lado $(v, w) = 1$, ya que si un primo p divide a ambos, entonces

$$\begin{aligned} p \quad &| \quad (v + w) = \frac{1}{2}(z + x) + \frac{1}{2}(z - x) = \frac{1}{2}\,2z = z, \\ p \quad &| \quad (v - w) = \frac{1}{2}(z + x) - \frac{1}{2}(z - x) = x, \end{aligned}$$

y como $(x, z) = 1$, esto es contradictorio.

Ahora usamos un argumento al que le sacaremos mucho partido:

> *Si el producto de dos números naturales primos entre sí es una potencia n-sima, entonces ambos factores son potencias n-simas.*

Esto se debe a que cada primo que divida a uno de los factores debe dividirlo con el mismo exponente que al producto, luego dicho exponente debe ser múltiplo de n.

En nuestro caso concluimos que v y w son cuadrados perfectos. Pongamos que $v = p^2$ y $w = q^2$. Obviamente $(p, q) = 1$. Así tenemos que

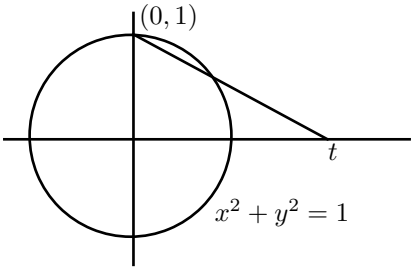
$$z = v + w = p^2 + q^2, \qquad x = v - w = p^2 - q^2.$$

En particular $q < p$. Como z y x son impares, p y q deben tener paridad opuesta. Sustituyendo en las fórmulas anteriores queda

$$y^2 = z^2 - x^2 = p^4 + 2p^2q^2 + q^4 - p^4 + 2p^2q^2 - q^4 = 4p^2q^2 = (2pq)^2,$$

luego $y = 2pq$ y la terna dada resulta ser de la forma indicada en el enunciado.

Recíprocamente, es fácil comprobar que cualquier terna en estas condiciones es una terna pitagórica primitiva. ∎



Vamos a dar una prueba geométrica alternativa del teorema 2.2. Para ello consideramos la proyección estereográfica, que a cada punto t de la recta numérica le asigna el punto de la circunferencia de ecuación $x^2 + y^2 = 1$ por donde pasa la recta que une t con el punto $(0, 1)$.

Los puntos de la recta son los que cumplen $x = (1 - y)t$ (pues ciertamente $(0, 1)$ y $(t, 0)$ cumplen la ecuación), luego el punto de corte con la circunferencia cumple

$$(1 - y)^2t^2 + y^2 = 1,$$

que equivale a

$$(1 + t^2)y^2 - 2t^2y + t^2 - 1 = 0,$$

o también:

$$y^2 - \frac{2t^2}{1 + t^2}y + \frac{t^2 - 1}{1 + t^2} = 0.$$

Este polinomio tiene dos raíces, una de las cuales es $y = 1$ y la otra es la coordenada y_0 del punto de intersección que buscamos. Por lo tanto, el polinomio factoriza como

$$(y - 1)(y - y_0) = y^2 - (y_0 + 1)y + y_0.$$

Por lo tanto,

$$y_0 = \frac{t^2 - 1}{t^2 + 1}.$$

De la ecuación de la recta sacamos el valor correspondiente de x_0 y concluimos que el punto de corte es

$$\left(\frac{2t}{t^2 + 1}, \frac{t^2 - 1}{t^2 + 1}\right).$$

Recíprocamente, si tenemos un punto (x_0, y_0) en la circunferencia distinto del punto $(0, 1)$, entonces la recta que pasa por ambos puntos tiene ecuación

$$(1 - y_0)x + x_0y - x_0 = 0,$$

luego corta al eje $y = 0$ en el punto $t = x_0/(1 - y_0)$. Tenemos así una correspondencia biunívoca entre los números racionales t y los puntos de la circunferencia unitaria distintos de $(0, 1)$ con coordenadas racionales.[1]

Todo lo anterior son hechos generales sobre la proyección estereográfica, que no sería justo considerar como parte de la demostración de 2.2 que vamos a ver a continuación. Es más razonable considerar que nuestra tercera prueba empieza aquí:

[1] Lo único que no era evidente es que la proyección estereográfica hace corresponder puntos racionales con puntos racionales.

DEMOSTRACIÓN: Supongamos que (x, y, z) es una terna pitagórica primitiva con y par. Entonces $(x/z)^2 + (y/z)^2 = 1$, luego $(y/z, x/z)$ es un punto de la circunferencia unitaria distinto de $(0, 1)$, luego, llamando

$$t = \frac{y/z}{1 - x/z} = \frac{y}{z - x},$$

hemos visto que se cumple

$$\frac{y}{z} = \frac{2t}{t^2 + 1}, \qquad \frac{x}{z} = \frac{t^2 - 1}{t^2 + 1}.$$

Si $t = p/q$, con p y q primos entre sí, entonces

$$\frac{y}{z} = \frac{2p/q}{p^2/q^2 + 1} = \frac{2pq}{p^2 + q^2}, \qquad \frac{x}{z} = \frac{(p/q)^2 - 1}{(p/q)^2 + 1} = \frac{p^2 - q^2}{p^2 + q^2}.$$

No puede ocurrir que p y q sean ambos impares, porque entonces p^2+q^2 sería par y en la primera fracción podríamos simplificar el 2, con lo que y sería impar. Entonces, la segunda fracción es irreducible,[2] porque un primo (necesariamente impar) que divida a su numerador y su denominador dividiría a p y a q, que son primos entre sí. Como y/z también es irreducible, tiene que ser $y = 2pq$, $z = p^2 + q^2$, luego la segunda fracción nos da que $x = p^2 - q^2$.

Así pues, llegamos a que toda terna pitagórica primitiva (con y par) es de la forma

$$(x, y, z) = (p^2 - q^2, 2pq, p^2 + q^2),$$

donde p y q son primos entre sí y tienen paridades opuestas. Además obviamente debe cumplirse que $q < p$ para que x sea positivo.

En estos términos es evidente que toda terna de esta forma es pitagórica, pues cumple que $(y/z, x/z)$ es el punto de la circunferencia unitaria que se corresponde con $t = p/q$, luego cumple $y^2/z^2 + x^2/z^2 = 1$, luego $x^2 + y^2 = z^2$. ∎

El Último Teorema de Fermat A continuación demostramos la parte del Último Teorema de Fermat que realmente probó Fermat, a saber, que la ecuación $x^4 + y^4 = z^4$ no tiene soluciones enteras positivas. En realidad probó algo ligeramente más general:

Teorema 2.3 *La ecuación $x^4 + y^4 = z^2$ no tiene soluciones enteras positivas. En particular el Último Teorema de Fermat es cierto para $n = 4$.*

DEMOSTRACIÓN: Si existen soluciones positivas de la ecuación $x^4+y^4 = z^2$, entonces (x^2, y^2, z) es una terna pitagórica. Notemos que si un primo p divide a x, y, z entonces $p^2 \mid z$ y $(x/p, y/p, z/p^2)$ sigue cumpliendo la ecuación. Aplicando

[2]Recordemos que una fracción de números enteros se dice *irreducible* si su numerador y su denominador son primos entre sí. Es fácil ver que todo número racional puede expresarse de forma única como una fracción irreducible.

esto un número finito de veces podemos suponer que $(x, y, z) = 1$, y claramente esto implica que en realidad son primos entre sí dos a dos y que la terna (x^2, y^2, z) es primitiva.

Según la clasificación de las ternas pitagóricas, $x^2 = p^2 - q^2$, $y^2 = 2pq$, $z = p^2 + q^2$, donde p y q son números enteros primos entre sí, de distinta paridad y $p > q > 0$ (intercambiamos x con y si es necesario para que y sea el número par).

Ahora, $p^2 = x^2 + q^2$, luego (q, x, p) es otra terna pitagórica, lo que obliga a que p sea impar, luego q ha de ser par, y así $q = 2ab$, $x = a^2 - b^2$, $p = a^2 + b^2$, para ciertos enteros a y b primos entre sí, de paridad opuesta, $a > b > 0$ (notemos que se trata de una terna primitiva porque $(p, q) = 1$).

Por lo tanto $y^2 = 4ab(a^2 + b^2)$ y en consecuencia $ab(a^2 + b^2) = (y/2)^2$. Por otra parte $(a, b) = 1$ implica fácilmente que $(ab, a^2 + b^2) = 1$.

Ahora volvemos a usar un argumento que ya hemos empleado al estudias las ternas pitagóricas:

Si el producto de dos números naturales primos entre sí es un cuadrado, entonces ambos son cuadrados, pues cada uno de ellos debe tener cada factor primo con exponente par.

Concluimos que ab y $a^2 + b^2$ son cuadrados y, por el mismo argumento, también lo son a y b. Digamos $a = u^2$, $b = v^2$, $a^2 + b^2 = w^2$. Entonces

$$u^4 + v^4 = a^2 + b^2 = w^2 = p < p^2 + q^2 = z < z^2.$$

En resumen, si existe una terna de números positivos (x, y, z) que satisfaga la ecuación $x^4 + y^4 = z^2$, existe otra (u, v, w) que cumple lo mismo pero con $w^2 < z^2$. Si existieran tales ternas debería haber una con z mínimo, lo cual es falso según lo visto, por lo que la ecuación no tiene solución. ∎

Es importante notar que el teorema anterior no sólo prueba el Último Teorema de Fermat para $n = 4$, sino en general para $n = 4k$. En efecto, si existieran números positivos (x, y, z) tales que $x^{4k} + y^{4k} = z^{4k}$, entonces (x^k, y^k, z^k) sería una solución a la ecuación $x^4 + y^4 = z^4$, lo cual es imposible. En particular el Último Teorema de Fermat es cierto para las potencias de dos.

De aquí se sigue ahora que si el Último teorema de Fermat es cierto para exponentes primos impares, entonces es cierto para todo exponente. En efecto, si existen soluciones positivas a una ecuación $x^n + y^n = z^n$, entonces n no puede ser potencia de 2, luego existe un primo impar p tal que $p \mid n$, o sea, $n = pk$, para cierto entero k, luego (x^k, y^k, z^k) es una solución positiva a la ecuación $x^p + y^p = z^p$.

Observemos también que si p es impar el Último Teorema de Fermat equivale a la no existencia de soluciones enteras no triviales (o sea, con $xyz \neq 0$) de la ecuación

$$x^p + y^p + z^p = 0,$$

lo que muestra que en realidad el papel de las tres variables es simétrico. Esto simplifica algunos argumentos.

Números de taxi En el capítulo IV estudiaremos los números que pueden expresarse como suma de dos cuadrados, pero aquí podemos decir algo (mucho más elemental) sobre los números que son suma de dos cubos. Es evidente que, para un n fijo, la ecuación $x^2 + y^2 = n$ sólo puede tener un número finito de soluciones enteras, pues necesariamente $|x|$, $|y| \leq \sqrt{n}$. En cambio, aunque no es difícil probarlo, no es inmediato que lo mismo vale para sumas de dos cubos:

Teorema 2.4 *Para cada entero $n \neq 0$, la ecuación $x^3 + y^3 = n$ tiene a lo sumo un número finito de soluciones enteras.*

DEMOSTRACIÓN: No sólo vamos a probar que sólo hay un número finito de soluciones, sino que vamos a ver cómo podemos obtenerlas todas. Nos basamos en la factorización:

$$x^3 + y^3 = (x + y)(x^2 - xy + y^2).$$

De ella deducimos que $x + y \mid n$, lo cual hace que sólo haya un número finito de posibilidades para $x + y$. Por otro lado,

$$\frac{n}{x+y} = x^2 - xy + y^2 = (x + y)^2 - 3xy,$$

luego

$$xy = \frac{1}{3}\left((x + y)^2 - \frac{n}{x+y}\right),$$

y por lo tanto, también hay un número finito de posibilidades para xy. Finalmente, x e y están unívocamente determinados por $b = x + y$ y $c = xy$, pues son las raíces del polinomio

$$(t - x)(t - y) = t^2 - (x + y)t + xy = t^2 - bt + c,$$

es decir,[3]

$$x, y = \frac{b \pm \sqrt{b^2 - 4c}}{2}.$$

Observemos que si $n > 0$, entonces $b > 0$, porque en caso contrario sería $x + y < 0$, $xy > 0$, lo que obligaría a que $x < 0$, $y < 0$, y entonces $n < 0$. ∎

Ejemplo Vamos a encontrar las soluciones enteras de la ecuación $x^3 + y^3 = 91$. La tabla siguiente contiene los cálculos necesarios:

b	c	D	x	y
1	−30	121	6	−5
7	12	1	4	3
13	54	−47		
91	2 760	−2 759		

[3]Usamos aquí la conocida fórmula para resolver ecuaciones de segundo grado, que discutiremos en el capítulo siguiente para poder precisar los cuerpos sobre los que puede aplicarse.

Aquí b recorre los divisores positivos de 91 y $c = \frac{1}{3}(b^2 - n/b)$. Si c es entero calculamos $D = b^2 - 4c$ y, si éste es cuadrado perfecto, calculamos x, y. Concluimos de este modo que 91 se expresa exactamente de dos formas distintas como suma de dos cubos:[4]

$$91 = 6^3 + (-5)^3 = 4^3 + 3^3.$$

Los números menores que 100 expresables como suma de dos cubos son los siguientes:

1	2	7	8	9	16	19
26	27	28	35	37	54	56
61	63	64	65	72	91	98

y, de entre todos ellos, el 91 es el único que puede expresarse como suma de dos cubos de dos formas distintas. Los siguientes son:

91	152	189	217	513	721	728
999	1 027	1 216	1 512	1 729	1 736	2 457

De entre todos ellos, 1 729 es el único que puede expresarse de dos formas distintas como suma de dos cubos positivos. La tabla siguiente muestra los cálculos correspondientes:

b	c	D	x	y
1	−576	2 305		
7	−66	313		
13	12	121	12	1
19	90	1	10	9
91	2 754	−2 735		
133	5 892	−5 879		
247	20 334	−20 327		
1 729	996 480	−996 479		

La conclusión es que

$$1\,729 = 12^3 + 1^3 = 10^3 + 9^3.$$

El siguiente número con esta propiedad es

$$4\,104 = 16^3 + 2^2 = 15^3 + 9^3,$$

que además admite la representación $4\,104 = 18^3 + (-12)^3$. Los números menores que 100 000 con esta propiedad son diez:

1 729	4 104	13 832	20 683	32 832
39 312	40 033	46 683	64 232	65 728

[4] Esta igualdad se expresa de forma más elegante como

$$3^3 + 4^3 + 5^3 = 6^3.$$

El menor número natural que puede expresarse como suma de dos cubos (positivos) de n formas distintas se conoce como el *número de taxi* (o de Hardy-Ramanujan) n-simo. El nombre se debe a una anécdota que contó Hardy sobre Ramanujan:

> *Recuerdo que una vez fui a verlo, cuando estaba enfermo en Putney. Había tomado el taxi número* 1 729*, y le dije que ese número me parecía bastante soso, y que esperaba que no fuera un mal presagio. No, —me respondió— es un número muy interesante. Es el menor número expresable como suma de dos cubos de dos formas distintas.*[5]

En palabras de Hardy: *"daba la impresión de que cada número natural era uno de sus amigos personales"*. En realidad la observación de Ramaujan la había publicado ya el matemático francés Bernard Frénicle de Bessy en 1657. El primer número de taxi es trivialmente $2 = 1^3 + 1^3$, mientras que el tercero fue calculado en 1957 con la ayuda de un ordenador. Resulta ser:

$$87\,539\,319 = 167^3 + 436^3 = 228^3 + 423^3 = 255^3 + 414^3.$$

Si el lector quiere encontrar este número por sus propios cálculos, le ayudará tener en cuenta que, para $x, y > 0$, necesariamente $\sqrt[3]{n} \le b \le \sqrt[3]{4n}$.

En efecto, $n/b = b^2 - 3c \le b^2$, luego $n \le b^3$ y, por otra parte, de la desigualdad $(x-y)^2 \ge 0$ se sigue que $xy \le x^2 - xy + y^2$, luego $c \le n/b$, luego $bc \le n$, de donde $b^3 = n + 3bc \le 4n$.

Ejercicio: Probar que 728 puede expresarse de tres formas distintas como suma de dos cubos (no necesariamente positivos). Es el menor número con esta propiedad.

Ahora vamos a probar lo siguiente:

Teorema 2.5 *La ecuación*

$$x^3 + y^3 = z^3 + w^3$$

tiene infinitas soluciones enteras no triviales, es decir, en las que $x \ne z$, $x \ne w$.

Demostración: Vamos a encontrar polinomios de segundo grado que nos proporcionen infinitas soluciones no triviales de esta ecuación. Nos basamos en que

$$\begin{aligned}(at^2 + bt + c)^3 &= a^3t^6 + 3a^2bt^5 + (3a^2c + 3ab^2)t^4 + (6abc + b^3)t^3 \\ &+ (3ac^2 + 3b^2c)t^2 + 3bc^2t + c^3\end{aligned}$$

[5]Hardy le preguntó entonces a Ramanujan cuál era el menor número natural expresable como suma de dos potencias cuartas de dos formas distintas, y Ramanujan le contestó que no lo sabía. Euler obtuvo la respuesta en 1772 mediante cálculos muy laboriosos. Hoy en día un ordenador obtiene la solución en segundos:

$$635\,318\,657 = 133^4 + 134^4 = 59^4 + 158^4.$$

Vamos a buscar dos polinomios distintos $p(t) = at^2 + bt + c$, $p'(t) = a't^2 + b't + c'$ tales que los coeficientes de las potencias impares de t en la expresión anterior sean opuestos. De este modo, el polinomio $f(t) = p(t)^3 + p'(t)^3$ tendrá sólo potencias pares de t, por lo que cumplirá $f(t) = f(-t)$ y tendremos la igualdad

$$(at^2 + bt + c)^3 + (a't^2 + b't + c')^3 = (at^2 - bt + c)^3 + (a't^2 - b't + c')^3,$$

que será no trivial si garantizamos además que $b' \neq \pm b$. Esto supone elegir los coeficientes de modo que cumplan las ecuaciones:

$$a^2b = -a'^2b', \qquad bc^2 = -b'c'^2, \qquad 6abc + b^3 = -6a'b'c' - b'^3.$$

No necesitamos encontrar todas las soluciones de este sistema, nos basta encontrar una, que sea lo más simple posible. Si probamos a hacer $a = 1$ queda

$$b = -a'^2b', \qquad bc^2 = -b'c'^2, \qquad 6bc + b^3 = -6a'b'c' - b'^3.$$

Ahora observamos que no podemos tomar $a' = \pm 1$, pues llegaríamos a $b = -b'$, y es justo lo que no queremos. Probamos entonces con $a' = 2$, y nos queda

$$b = -4b', \qquad bc^2 = -b'c'^2, \qquad 6bc + b^3 = -12b'c' - b'^3.$$

Sustituyendo la primera ecuación en las otras queda

$$b = -4b', \qquad 4c^2 = c'^2, \qquad -24c - 64b'^2 = -12c' - b'^2.$$

Por lo tanto tiene que ser $c' = \pm 2c$, luego

$$b = 4b', \qquad c' = \pm 2c, \qquad -24c - 64b'^2 = \mp 24c - b'^2.$$

La tercera ecuación no tiene solución con el signo negativo, así que tiene que ser

$$b = 4b', \qquad c' = -2c, \qquad -24c - 64b'^2 = 24c - b'^2.$$

La última ecuación equivale a $63b'^2 = -48c$, o también a $-21b'^2 = 2^4c$, con lo que la solución más simple es $c = -21$, $b' = 4$. En total hemos obtenido $(a, b, c) = (1, -16, -21)$, $(a', b', c') = (2, 4, 42)$, de donde resulta la identidad:

$$(t^2 - 16t - 21)^3 + (2t^2 + 4t + 42)^3 = (t^2 + 16t - 21)^3 + (2t^2 - 4t + 42)^3.$$

Ambos miembros son el polinomio

$$7t^6 - 105t^4 + 27\,405t^2 + 83\,349.$$

Es inmediato que la igualdad $t^2 - 16t - 21 = t^2 + 16t - 21$ solo se da cuando $t = 0$, mientras que la ecuación $t^2 - 16t - 21 = 2t^2 - 4t + 42$ equivale a $t^2 + 12t + 63 = (t + 6)^2 + 27 = 0$, que obviamente no tiene soluciones enteras. Por lo tanto, dando a t valores enteros positivos obtenemos infinitas soluciones no triviales de la ecuación inicial (son todas distintas entre sí porque el polinomio de sexto grado que hemos obtenido, al tener todos sus coeficientes positivos, proporciona obviamente números mayores a medida que aumenta t). ∎

Observemos que el primer ejemplo explícito de solución que proporciona el argumento del teorema anterior, es decir, el que resulta de hacer $t = 1$, es

$$(-36)^3 + 48^3 = (-4)^3 + 40^3.$$

Ahora bien, todos los números son múltiplos de 4, luego si dividimos la igualdad entre 4^3 obtenemos

$$-9^3 + 12^3 = -1^3 + 10^3,$$

que es la solución correspondiente al "número de taxi", pero desordenada. Un poco más en general, si cambiamos t por $2t + 1$, obtenemos:

$$(4t^2 - 28t - 36)^3 + (8t^2 + 16t + 48)^3 = (4t^2 + 36t - 4)^3 + (8t^2 + 40)^3.$$

Vemos que todos los coeficientes son múltiplos de 4, luego podemos dividir la igualdad entre 4^3 y obtenemos otra expresión que nos da soluciones "más pequeñas". Si además la reordenamos conseguimos que la correspondiente a $t = 0$ sea precisamente el número de taxi:

$$(2t^2 + 4t + 12)^3 + (-t^2 - 9t + 1)^3 = (2t^2 + 10)^3 + (-t^2 + 7t + 9)^3.$$

El número de divisores En 1644 Mersenne planteó el problema de encontrar el menor número con exactamente 60 divisores (positivos). Vamos a presentar algunos resultados que nos permitirán resolver este problema y otros similares.

Es costumbre llamar $d(n)$ al número de divisores (positivos) de un número natural n. Si $n = p_1^{e_1} \cdots p_r^{e_r}$, es muy fácil concluir que

$$d(n) = (e_1 + 1) \cdots (e_r + 1).$$

Por ejemplo, $d(1\,000\,000) = d(2^6 \cdot 5^6) = (6 + 1)(6 + 1) = 49$.

Resolveremos el problema de Mersenne a partir de un hecho elemental:

Teorema 2.6 *Si d, $d' > 1$ y $p^d > q$, entonces $p^{dd'-1} > p^{d-1}q^{d-1}$.*

DEMOSTRACIÓN: Como $d' > 1$, tenemos que $d' - 1 > 0$, luego se cumple $p^{d(d'-1)} > q^{d'-1}$ y, multiplicando por p^{d-1}, obtenemos $p^{dd'-1} > p^{d-1}q^{d'-1}$. ∎

A su vez:

Teorema 2.7 *Si $M(n)$ es el menor número natural con exactamente n divisores y $p_1, p_2, p_3, \ldots$ es la sucesión de los números primos, entonces*

$$M(n) = p_1^{e_1-1} p_2^{e_2-1} \cdots p_k^{e_k-1},$$

donde los exponentes cumplen $e_1 \geq e_2 \geq \cdots \geq e_k \geq 2$, $e_1 \cdots e_k = n$ y, si $d \mid e_i$ es un divisor propio (distinto de 1 y e_i), entonces $p_i^d < p_{k+1}$.

DEMOSTRACIÓN: Los divisores primos de $M(n)$ tienen que ser los primeros primos $p_1, \ldots, p_k$ hasta un cierto k, pues si hubiera un "hueco", es decir, si $p_i \nmid M(n)$, $p_j \mid M(n)$, con $i < j$, el número que resulta de sustituir $p_j^{e_j-1}$ por $p_i^{e_j-1}$ en la factorización de $M(n)$ sería estrictamente menor y tendría el mismo número de divisores primos. Además, si $i < j$ entonces $e_i \geq e_j$, pues en caso contrario, si $e_i < e_j$, se cumple que $p_i^{e_i-1}p_j^{e_j-1} > p_i^{e_j-1}p_j^{e_i-1}$ (pues estamos sustituyendo $e_j - e_i$ primos p_j por primos menores p_i), luego cambiando estas potencias en la factorización de $M(n)$ obtenemos un número menor con el mismo número de divisores. Además $n = d(M(n)) = e_1 \cdots e_k$.

Finalmente, si $e_i = dd'$, con d, $d' > 1$, pero $p_i^d \geq p_{k+1}$, como las bases son primos distintos, tiene que ser $p_i^d > p_{k+1}$, y el teorema anterior nos da que $p_i^{e_i-1} > p_i^{d-1}p_{k+1}^{d'-1}$, luego cambiando $p_i^{e_i-1}$ por $p_i^{d-1}p_{k+1}^{d'-1}$ en la factorización de $M(n)$ obtenemos un número menor con el mismo número de divisores. ∎

Si q es primo, el teorema anterior implica que $e_1 = q$, por lo que $M(q) = 2^{q-1}$.

Ahora, si $n = q_1q_2$, para dos primos $q_1 \geq q_2$, entonces $M(n) = 2^{q_1-1}3^{q_2-1}$.

En efecto, el teorema anterior deja en principio dos alternativas. Una es la que $e_1 = q_1$, $e_2 = q_2$, que lleva a la conclusión que queremos probar, y la otra es que $e_1 = q_1q_2$, en cuyo caso $M(n) = 2^{q_1q_2-1} > 2^{q_1-1}3^{q_2-1}$, pero entonces, tomando $d = q_1$, tendríamos que $2^{q_1} < p_2 = 3$, lo cual es imposible.

Supongamos finalmente que $n = q_1q_2q_3$ o $n = q_1q_2q_3q_4$, donde los primos q_i forman una sucesión decreciente. Entonces, si $e_1 = dd'$, con $d \geq 3$ y $d' \geq 2$, entonces $k \leq 3$, luego

$$2^d \geq 2^3 > 7 = p_4 \geq p_{k+1},$$

en contradicción con el teorema anterior. Esto implica que e_1 es primo o $e_1 = 4$.

Por otro lado, si $i \geq 2$ y $e_i = dd'$, con d, $d' \geq 2$, entonces

$$p_i^d \geq p_i^2 \geq 3^2 > 7 = p_4 \geq p_{k+1},$$

y de nuevo tenemos una contradicción, luego e_i es primo.

Así pues, si $n = q_1q_2q_3$, o bien $e_1 = 4$, $e_2 = q_1$, en cuyo caso $n = 4q_1$ y $M(n) = 2^{4-1} \cdot 3^{q_1-1}$, con $e_1 = 4 \geq e_2 = q_1 \geq 2$ o, más precisamente, dado que q_1 es primo, $2 \leq q_1 \leq 3$, o bien tenemos $e_1 = q_1$, $e_2 = q_2$, $e_3 = q_3$, luego $M(n) = 2^{q_1-1}3^{q_2-1}5^{q_3-1}$.

El primer caso sólo puede darse si $n = 8$ o $n = 12$, pero entonces:

q_1	n	$2^{4-1} \cdot 3^{q_1-1}$	$2^{q_1-1}3^{q_2-1}5^{q_3-1}$
2	8	24	30
3	12	72	60

por lo que el primer caso sólo se da con $M(8) = 24$.

Por otro lado, si $n = q_1q_2q_3q_4$, o bien $e_1 = 4$, $e_2 = q_1$, $e_3 = q_2$, en cuyo caso $n = 4q_1q_2$, con $e_1 = 4 \geq e_2 = q_1 \geq e_3 = q_2 \geq 2$, luego de hecho $2 \leq q_2 \leq q_1 \leq 3$ y $M(n) = 2^{4-1} \cdot 3^{q_1-1} \cdot 5^{q_2-1}$, o bien todos los e_i son primos y se cumple que $M(n) = 2^{q_1-1} \cdot 3^{q_2-1} \cdot 5^{q_3-1} \cdot 7^{q_4-1}$.

El primer caso sólo puede darse para los valores de la tabla siguiente:

q_1	q_2	n	$2^{4-1}\cdot 3^{q_1-1}\cdot 5^{q_2-1}$	$2^{q_1-1}\cdot 3^{q_2-1}\cdot 5^{q_3-1}\cdot 7^{q_4-1}$
2	2	16	120	210
3	2	24	360	420
3	3	36	1 800	1 260

vemos que hay que descartar el 36 y la conclusión es la siguiente:

Teorema 2.8 *Si $n = q_1 \cdots q_k$ con $k \leq 4$ y los primos q_i forman una sucesión decreciente, entonces el menor número con n divisores es $M(n) = p_1^{q_1-1}\cdots p_k^{q_k-1}$ excepto en los casos $n = 8, 16, 24$, para los cuales se cumple que $M(8) = 24$, $M(16) = 120$ y $M(24) = 360$.*

En particular hemos resuelto el problema de Mersenne:

Ejemplo: *El menor número con 60 divisores es $M(60) = 2^4\cdot 3^2\cdot 5\cdot 7 = 5\,040$.*

Basta aplicar el teorema anterior teniendo en cuenta que $60 = 5\cdot 3\cdot 2\cdot 2$.

Ejercicio: Calcular el menor número con 100 divisores.

Notemos que $M(2) = 2$, pero

$$M(3) = 4, \quad M(4) = 6, \quad M(6) = 12, \quad M(12) = 60, \quad M(60) = 5\,040.$$

Esto nos lleva a:

Ejemplo El menor número con 5 040 divisores es 293 318 625 600.

En efecto, como $5\,040 = 2^4\cdot 3^2\cdot 5\cdot 7$ tiene 8 factores primos, el teorema 2.7 nos da que $M(5\,040)$ es divisible a lo sumo entre 8 primos distintos, por lo que $p_{k+1} \leq 23$. Por lo tanto, si d es un divisor propio de e_i, con $i \geq 3$, tiene que ser $5^2 \leq p_i^d < 23$, lo cual es imposible, luego e_i es primo salvo a lo sumo si $i = 1, 2$.

El mismo razonamiento prueba que un divisor propio de e_1 tiene que ser menor o igual que 4, luego, si e_1 no es primo, siendo divisor de 5 040, tiene que ser uno de los valores $4, 6, 8, 9, 12$, pero tenemos que descartar el 4 y el 6 porque entonces 7 dividiría a otro e_i y no se cumpliría $e_1 \geq e_i$. Por otro lado, si e_1 es primo tiene que ser $e_1 = 7$.

Similarmente, un divisor propio de e_2 tiene que ser menor o igual que 2, luego si e_2 no es primo, tiene que ser 4, pero esto es imposible, porque, como 5 no puede dividir a e_1, tendría que ser un e_i, y no se cumpliría $e_2 \geq e_i$. Así pues, todos los e_i son primos salvo quizá e_1.

La tabla siguiente recoge todas las posibilidades:

e_1	M	
12	$2^{11}\cdot 3^6\cdot 5^4\cdot 7^2\cdot 11\cdot 13$	6 538 371 840 000
9	$2^8\cdot 3^6\cdot 5^4\cdot 7\cdot 11\cdot 13\cdot 17$	1 984 862 880 000
8	$2^7\cdot 3^6\cdot 5^4\cdot 7^2\cdot 11^2\cdot 13$	4 495 130 640 000
7	$2^6\cdot 3^4\cdot 5^2\cdot 7^2\cdot 11\cdot 13\cdot 17\cdot 19$	293 318 625 600

Por lo tanto, el valor de $M(5\,040)$ es el correspondiente a la última fila. ∎

Vamos a ver que los números $M(q^n)$, donde q es primo, son fáciles de calcular. Para ello, si p_i es el primo i-ésimo, definimos

$$G(i,l) = p_i^{(q-1)q^{l-1}}, \qquad i,l \geq 1,$$

y llamamos $g_1, g_2, \ldots$ a la sucesión que forman estos números en orden creciente. Por ejemplo, para $q = 2$ tenemos que

$$\begin{array}{cccccc} 2 & 4 & 8 & 16 & 32 & 64 \\ 3 & 9 & 27 & 81 & \cdots & \\ 5 & 25 & \cdots & & & \\ 7 & 49 & \cdots & & & \\ 11 & \cdots & & & & \end{array}$$

donde dejamos de calcular cuando llegamos a números mayores que 100. La primera columna se prolonga con la sucesión de los primos, por los términos menores que 100 de la sucesión son

$$\begin{array}{cccccccccccccccc} 2 & 3 & 4 & 5 & 7 & 9 & 11 & 13 & 16 & 17 & 19 & 23 & 25 & 29 & 31 & 37 \\ 41 & 43 & 47 & 49 & 53 & 59 & 61 & 67 & 71 & 73 & 79 & 81 & 83 & 89 & 97 & \cdots \end{array}$$

Dejamos al lector la comprobación de que, para $q = 3$, los términos menores que 1 000 de la sucesión g_i son

$$4 \quad 9 \quad 25 \quad 49 \quad 64 \quad 121 \quad 169 \quad 289 \quad 361 \quad 529 \quad 729 \quad 841 \quad 961 \quad \cdots$$

Teorema 2.9 *En las condiciones anteriores, se cumple que* $M(q^m) = g_1 \cdots g_m$.

Por ejemplo, el menor número con 3^{10} divisores es

$$M(3^{10}) = 4 \cdot 9 \cdot 25 \cdot 49 \cdot 64 \cdot 121 \cdot 169 \cdot 289 \cdot 361 \cdot 529 = 3\,185\,307\,433\,269\,561\,600.$$

Ejercicio: Calcular el menor número con 2^{10} divisores.

Demostración: Sea n un número con $d(n) = q^m$ divisores. Pongamos que su descomposición en factores primos es $n = q_1^{e_1} \cdots q_r^{e_r}$. Entonces se cumple que $d(n) = (e_1 + 1) \cdots (e_r + 1) = q^m$, luego $e_i + 1 = q^{h_i}$ con $h_1 + \cdots + h_r = m$. Así pues, $n = q_1^{q^{h_1}-1} \cdots q_r^{q^{h_r}-1}$. Ahora observamos que

$$q_i^{q^{h_i}-1} = q_i^{(q-1)(1+q+\cdots+q^{h_i-1})} = q_i^{q-1} q_i^{(q-1)q} \cdots q_i^{(q-1)q^{h_i-1}}.$$

Por lo tanto, $q_i^{q^{h_i}-1}$ es producto de h_i números g_j distintos entre sí, todos con base q_i, luego son distintos de los correspondientes a cualquier otro primo $q_{i'}$, y en total n es producto de $h_1 + \cdots + h_r = m$ números g_j distintos entre sí. Por lo tanto, $n \geq g_1 \cdots g_m$. Ahora basta probar que

$$d(g_1 \cdots g_m) = q^m.$$

Razonamos por inducción sobre m. Si $m = 1$ tenemos que $g_1 = G(1,1) = 2^{q-1}$, luego $d(g_1) = q$. Si es cierto para m y $g_{m+1} = G(i,l)$, entonces todos los $G(i,l')$ con $l' < l$ aparecen entre $g_1, \dots, g_m$, luego el exponente de p_i en $g_1 \cdots g_m$ es $(q-1)(1+q+\cdots+q^{l-2}) = q^{l-1}-1$ y el exponente en $g_1 \cdots g_{m+1}$ es $q^l - 1$, luego

$$d(g_1 \cdots g_{m+1}) = d(g_1 \cdots g_m)\frac{q^l}{q^{l-1}} = q^m q = q^{m+1}.$$

■

2.3 Dominios de factorización única

Vamos a ver cómo se adaptan los conceptos aritméticos básicos cuando se aplican a $\mathbb{Z}$ en lugar de a $\mathbb{N}$. Vamos a ver que los conceptos aritméticos "enturbian" un poco al considerar números enteros, pero que esto se compensa con creces con la posibilidad de restar sin restricciones que ofrece el anillo de los números enteros. Además, las adaptaciones necesarias para trabajar en $\mathbb{Z}$ nos permiten en realidad extender los resultados aritméticos básicos a una clase relativamente amplia de anillos, entre ellos los anillos de polinomios, como $\mathbb{Z}[x]$ o $\mathbb{Q}[x]$, de modo que el caso de $\mathbb{Z}$ resulta ser un mero caso particular.

Para empezar, el concepto de divisibilidad que hemos definido en $\mathbb{N}$ es válido igualmente en cualquier dominio íntegro:

Definición 2.10 Si a y b son dos elementos de un dominio íntegro A, diremos que a *divide* a b, (o que b es un *múltiplo* de a o que b es *divisible* entre a), y lo representaremos por $a \mid b$, si existe un elemento c de A tal que $b = ac$. En caso contrario escribiremos $a \nmid b$.

En el caso concreto de $\mathbb{Z}$ es fácil ver que dos números enteros m y n cumplen $m \mid n$ si y sólo si $|m| \mid |n|$ en el sentido de la sección precedente. Así, por ejemplo, si los divisores de 6 en $\mathbb{N}$ son $1, 2, 3, 6$, sus divisores en $\mathbb{Z}$ son los números enteros cuyo valor absoluto divide a 6, es decir, $\pm 1, \pm 2, \pm 3, \pm 6$. Vemos que al pasar de $\mathbb{N}$ a $\mathbb{Z}$, los divisores de un número "se duplican".

Es inmediato que la divisibilidad cumple las propiedades:

Reflexiva $a \mid a$,

Transitiva Si $a \mid b$ y $b \mid c$, entonces $a \mid c$.

Sin embargo, mientras que la divisibilidad en $\mathbb{N}$ es antisimétrica, es decir, si dos números naturales cumplen $m \mid n$ y $n \mid m$, necesariamente $m = n$, en $\mathbb{Z}$ tenemos, por ejemplo, que $6 \mid -6$ y $-6 \mid 6$, pero obviamente $6 \neq -6$.

Elementos asociados Nos acabamos de encontrar con que en $\mathbb{Z}$ se da un fenómeno que no aparece cuando trabajamos exclusivamente con números naturales: la existencia de números asociados.

Definición 2.11 Diremos que dos elementos a y b de un dominio íntegro A son *asociados* en A, y lo representaremos por $a \sim b$, si $a \mid b$ y $b \mid a$.

Es claro que la relación de asociación cumple las propiedades siguientes:

Reflexiva $a \sim a$.

Simétrica *Si* $a \sim b$, *entonces* $b \sim a$.

Transitiva *Si* $a \sim b$ *y* $b \sim c$, *entonces* $a \sim c$.

Más aún, es inmediato que *dos elementos son asociados si y sólo si tienen los mismos múltiplos y los mismos divisores.*

En general, veremos que, en la práctica, que dos elementos sean asociados se traduce en que son "la misma cosa" a efectos de divisibilidad.

Por ejemplo, en el caso de $\mathbb{Z}$, la asociación agrupa a los números enteros en parejas (excepto en el caso del cero):

$$0, \quad \pm 1, \quad \pm 2, \quad \pm 3, \quad \ldots$$

En otros términos: dos números enteros cumplen $m \sim n$ si y sólo si $|m| = |n|$.

El caso del 0 es un caso aparte pues es fácil ver que en un dominio íntegro arbitrario el 0 es el único asociado de sí mismo. En los demás casos, la distribución de los asociados "por parejas" no es casual. En genera, la presencia de elementos asociados en un dominio íntegro es consecuencia de la existencia de unidades distintas de 1. Recordemos que una unidad es un elemento tal que exista u^{-1} de modo que $uu^{-1} = 1$.

Equivalentemente, las unidades son los divisores de 1 y, como 1 divide a cualquier otro elemento, podemos concluir que las unidades de un dominio íntegro son los asociados a 1. Visto así, el teorema siguiente afirma que los asociados a 1 determinan los asociados de cualquier otro elemento:

Teorema 2.12 *Dos elementos a y b de un dominio íntegro A son asociados si y sólo si existe una unidad u de A tal que $b = ua$.*

Demostración: Si u es una unidad, entonces claramente $a \mid ua$ y también $ua \mid a$, pues $a = u^{-1}(ua)$, luego $a \sim ua$.

Recíprocamente, si $a \sim b$, entonces existen u y v tales que $b = ua$ y $a = vb$, luego $b = uvb$. Si $b = 0$ tenemos que $a = v0 = 0$, luego obviamente $a = b \cdot 1$ y se cumple lo requerido con $u = 1$. Si $b \neq 0$ podemos simplificarlo y concluir que $uv = 1$, luego u es una unidad. ∎

Así, como las unidades de $\mathbb{Z}$ son ± 1, resulta que cada número entero m no nulo tiene exactamente dos asociados, que son $m \cdot 1$ y $m \cdot (-1)$, es decir, $\pm m$.

Elementos irreducibles y primos Hemos definido un número natural primo como un número $p \geq 2$ cuyos únicos divisores son 1 y p, pero si queremos adaptar la definición para adecuarla a los números enteros debemos tener en cuenta que, por ejemplo, 5 ya no tiene únicamente divisores 1 y 5, sino que sus divisores son ± 1 y ± 5. Si pensamos en un dominio íntegro arbitrario, debemos tener en cuenta que todo elemento a es divisible como mínimo entre sus asociados y entre las unidades. De hecho, en una descomposición $a = bc$, se cumple que b es una unidad si y sólo si c es un asociado de a y viceversa.

Esto nos lleva a las definiciones siguientes:

Definición 2.13 Un elemento p de un dominio íntegro A es *irreducible* si no es 0 ni una unidad y sus únicos divisores son las unidades de A y sus asociados en A.

Se dice que A es un *dominio de factorización única* si todo elemento de A que no sea nulo ni una unidad se descompone en producto de factores irreducibles de forma única salvo orden o asociación.

Por ejemplo, es fácil ver que

$$110 = 2 \cdot 5 \cdot 11 = (-5) \cdot 2 \cdot (-11)$$

son dos descomposiciones en factores irreducibles de 110 en $\mathbb{Z}$, y no son la misma, pero se cumple que $2 \sim 2$, $5 \sim -5$ y $11 \sim -11$.

Así, la unicidad salvo orden y asociación quiere decir en general que si tenemos dos descomposiciones

$$a = p_1 \cdots p_r = q_1 \cdots q_s$$

en factores irreducibles, necesariamente $r = s$ y, reordenando si es preciso los factores, se cumple que cada $p_i \sim q_i$.

Notemos que $\mathbb{N}$ no es un dominio de factorización única simplemente porque no es un dominio. Sin embargo, el teorema fundamental de la aritmética se traduce fácilmente en que:

$\mathbb{Z}$ es un dominio de factorización única.

En efecto, en primer lugar observemos que un número entero p es irreducible en $\mathbb{Z}$ si y sólo si el número natural $|p|$ es primo en el sentido de la sección anterior. En efecto, ambas condiciones implican que $p \neq 0$ y $p \neq \pm 1$ y la segunda equivale a que los divisores de $|p|$ sean 1 y p, lo cual equivale a que los divisores de p en $\mathbb{Z}$ sean ± 1 y $\pm p$, es decir, las dos unidades y los dos asociados de p, y esto es tanto como decir que p es irreducible.

En suma, los enteros irreducibles son $\pm 2, \pm 3, \pm 5, \pm 7, \ldots$

Por otro lado, dado un entero n no nulo ni unitario, una descomposición en factores primos de $|n| = p_1 \cdots p_r$ da lugar a una descomposición en factores

irreducibles de n sin más que cambiar p_1 por $-p_1$ en el caso de que n sea negativo. Y si tenemos dos descomposiciones en factores irreducibles

$$p_1 \cdots p_r = q_1 \cdots q_s,$$

entonces $|p_1| \cdots |p_r| = |q_1| \cdots |q_s|$ son dos descomposiciones en factores primos de un mismo número natural, luego por el teorema fundamental de la aritmética $r = s$ y, reordenado los factores si es preciso, se cumple que $|p_i| = |q_i|$, es decir, que $p_i \sim q_i$. Por lo tanto, las dos descomposiciones son la misma salvo orden y asociación.

La razón por la que hemos llamado "elementos irreducibles" a lo que hubiera sido natural llamar "elementos primos" es que es costumbre reservar este nombre para elementos con una propiedad más fuerte:

Definición 2.14 Un elemento p de un dominio íntegro A es *primo* si no es 0 ni una unidad y, para todo par de elementos a, b de A, se cumple que si $p \mid ab$, entonces $p \mid a$ o $p \mid b$.

La relación entre los elementos irreducibles y los primos es la siguiente:

Teorema 2.15 *En un dominio íntegro, todo elemento primo es irreducible. En un dominio de factorización única, todo elemento irreducible es primo.*

Demostración: Si p es primo, entonces no es nulo ni unitario. Veamos que los únicos divisores de p son unidades o asociados. En efecto, si $a \mid p$, existe un b tal que $p = ab$. Como $p \mid ab$, o bien $p \mid a$ o bien $p \mid b$. Pero a y b dividen a p, luego $a \sim p$ o $b \sim p$. En el segundo caso a es una unidad.

Supongamos ahora que p es irreducible en un dominio de factorización única. Por definición no es nulo ni una unidad. Supongamos ahora que $p \mid ab$, pero que $p \nmid a$ y $p \nmid b$. Si a es una unidad, entonces $p \mid b$. Igualmente si b es una unidad se cumple que $p \mid a$. Supongamos, pues, que a y b no son unidades. Entonces, en una descomposición en factores irreducibles de a y de b, no puede aparecer ningún factor asociado a p. Al multiplicar ambas descomposiciones obtenemos una descomposición en factores irreducibles de ab en las que ningún factor es asociado a p.

Por otra parte, si $ab = pc$, al multiplicar por p una descomposición en factores irreducibles de c obtenemos una descomposición en factores irreducibles de ab en la que sí que aparece un factor igual a p. Esto contradice la unicidad de la factorización. ∎

Así pues, al trabajar en dominios de factorización única podemos hablar de "elementos primos" en lugar de "elementos irreducibles", pues ambos conceptos son equivalentes.

Hemos visto que la única diferencia a la hora de descomponer un número natural en factores primos en $\mathbb{N}$ o en $\mathbb{Z}$ es que en $\mathbb{N}$ los factores son necesariamente positivos, mientras que en $\mathbb{Z}$ pueden ser positivos o negativos, y no hay

ninguna diferencia estrictamente aritmética entre cada primo positivo y su asociado negativo. Sin embargo, hay una precaución que debemos tener presente cuando trabajamos con descomposiciones en factores primos en $\mathbb{Z}$ o en dominios de factorización única arbitrarios:

En el caso de $\mathbb{N}$ es evidente que podemos agrupar en potencias los factores primos repetidos, de modo que todo número natural $n \geq 2$ admite una descomposición en factores primos de la forma

$$n = p_1^{e_1} \cdots p_r^{e_r},$$

donde los primos $p_1, \ldots, p_r$ son distintos dos a dos. Sin embargo, esto ya no es cierto en $\mathbb{Z}$. Por ejemplo, es imposible expresar de esta forma el número -100. Una descomposición en factores primos es $-100 = -2 \cdot 2 \cdot 5 \cdot 5$, pero 2 y -2 son primos asociados que no podemos agrupar, a no ser que dejemos fuera el -1, así:

$$-100 = -2^2 \cdot 5^2.$$

La situación general es la siguiente:

> *Todo elemento a de un dominio de factorización única que no sea nulo ni unitario puede descomponerse en la forma*
>
> $$a = u\, p_1^{e_1} \cdots p_r^{e_r},$$
>
> *donde los elementos $p_1, \ldots, p_r$ son primos no asociados dos a dos, los exponentes e_i son números naturales no nulos y u es una unidad.*

En efecto, lo que nos garantiza la definición de dominio de factorización única es que a puede descomponerse en la forma

$$a = q_1 \cdots q_s,$$

donde los q_i son primos. Ahora, si los reordenamos de modo que $q_1, \ldots q_{e_1}$ sean asociados entre sí, para agruparlos en forma de potencia tenemos que expresarlos en la forma $q_i = v_i p_1$, donde v_i es una unidad, de modo que

$$q_1 \cdots q_{e_1} = v_1 \cdots v_{e_1} p^{e_1} = u_1 p^{e_1},$$

donde $u_1 = v_1 \cdots v_{e_1}$ es una unidad.[6]

Si vamos agrupando de este modo los grupos de factores primos asociados entre sí, la descomposición de a quedará de la forma

$$a = u_1 \cdots u_r p_1^{e_1} \cdots p_r^{e_r} = u p_1^{e_1} \cdots p_r^{e_r},$$

pero el ejemplo precedente muestra que en general la unidad u no puede eliminarse. La unicidad de la factorización se traduce en que los exponentes e_i están unívocamente determinados.

[6]Notemos que, en general, el producto de unidades es una unidad.

Definición 2.16 Si a es un elemento no nulo ni unitario de un dominio de factorización única A y p es un primo, llamaremos *valor p-ádico* $v_p(a)$ de a al número de factores primos asociados a p que aparecen en una descomposición es factores primos de a. Si a es una unidad, definimos $v_p(a) = 0$.

Así, si $a = up_1^{e_1} \cdots p_r^{e_r}$ es la descomposición en factores primos de a (donde los p_i son no asociados dos a dos), entonces $v_{p_i}(a) = e_i$, mientras que $v_p(a) = 0$ para todo primo p que no sea asociado a ninguno de los p_i.

Por ejemplo, el hecho de que $500 = 2^2 \cdot 5^3$ es casi equivalente a que se cumple que $v_2(500) = 2$, $v_5(500) = 3$ y $v_p(500) = 0$ para todo entero primo p no asociado a 2 ni a 5. El "casi" se debe a que -500 cumple lo mismo. Los valores p-ádicos, como todos los conceptos relacionados con la divisibilidad que vamos a introducir, no distinguen entre un elemento y sus asociados.

Veamos algunas propiedades elementales de los valores p-ádicos. En ellas p representa un elemento irredudible de un dominio de factorización única A y los elementos a, b considerados son no nulos.[7]

1. *a es una unidad si y sólo si $v_p(a) = 0$ para todo p.*

   Si a es una unidad, se cumple que $v_p(a) = 0$ por definición del valor p-ádico. Si a no es una unidad, entonces admite una descomposición en factores primos, y si p es cualquiera de los factores que aparecen en ella, se cumple que $v_p(a) \geq 1$.

2. $v_p(ab) = v_p(a) + v_p(b)$.

   Si a y b son ambos unidades, se sigue de la propiedad precedente. Si a no es una unidad pero b sí que lo es, tomamos una descomposición en factores primos $a = up_1^{e_1} \cdots p_r^{e_r}$ y tenemos que $ab = bup_1^{e_1} \cdots p_r^{e_r}$ es una descomposición en factores primos de ab, luego se cumple que $v_p(ab) = v_p(a) = v_p(a) + v_p(b)$. Igualmente se razona si a es una unidad y b no.

   Si a y b no son unidades, una descomposición en factores primos de ab se obtiene multiplicando una de a por otra de b y, si en éstas hay $v_p(a)$ y $v_p(b)$ factores asociados a p, respectivamente, en la obtenida al multiplicarlas habrá $v_p(a) + v_p(b)$ factores.

3. *$p^n \mid a$ si y sólo si $v_p(a) \geq n$.* (Esto se expresa a menudo diciendo que $v_p(a)$ es el número de veces que p divide a a.)

   En efecto, si $p^n \mid a$, entonces existe un b tal que $a = p^n b$ y, descomponiendo b en factores primos, obtenemos una descomposición en factores primos de a que contiene al menos n factores iguales a p, luego $v_p(a) \geq n$.

   Recíprocamente, si $v_p(a) \geq n$, entonces $a = up_1^{e_1} \cdots p_r^{e_r}$, donde existe un i tal que $p_i \sim p$ y $e_i \geq n$. Entonces $p_i = vp$, para cierta unidad v, luego $p^n \mid p_i^{e_i} \mid a$.

[7] Conviene observar que no es necesario exceptuar el 0 en ningún caso si convenimos en que $v_p(0) = +\infty$, así como que $+\infty$ es mayor que todo número natural y que $+\infty + n = +\infty + \infty = +\infty$.

4. $a \mid b$ *si y sólo si* $v_p(a) \leq v_p(b)$ *para todo* p.

   Si $a \mid b$, existe un c tal que $b = ac$, luego $v_p(a) \leq v_p(a) + v_p(c) = v_p(b)$.

   Si se da esta condición, podemos suponer que a no es una unidad, pues en caso contrario $a \mid b$ trivialmente. Sea $a = up_1^{e_1} \cdots p_r^{e_r}$. Por hipótesis $e_i = v_{p_i}(a) \leq v_{p_i}(b)$, luego b admite una descomposición de la forma $b = p_1'^{e_1'} \cdots p_r'^{e_r'} c$, donde $p_i' \sim p_i$, $e_i' \geq e_i$ y c es una unidad o un producto de de primos no asociados a $p_1', \ldots, p_r'$. Ahora bien, expresando $p_i' = u_i p_i$ y agrupando cada unidad $u_i^{e_i'}$ con c, podemos suponer que $p_i' = p_i$, y entonces $b = au^{-1} p_1^{e_1'-e_1} \cdots p_r^{e_r'-e_r} c$, luego $a \mid b$.

5. $a \sim b$ *si y sólo si* $v_p(a) = v_p(b)$ *para todo* p.

   Es consecuencia inmediata de la propiedad precedente.

6. *Si* $a, b, a+b \neq 0$, *entonces* $v_p(a+b) \geq \min\{v_p(a), v_p(b)\}$.

   Si además $v_p(a) \neq v_p(b)$, *entonces* $v_p(a+b) = \min\{v_p(a), v_p(b)\}$.

   Podemos suponer sin pérdida de generalidad que $v_p(a) = e \leq e' = v_p(b)$. Entonces $a = p^e a'$, $b = p^{e'} b'$, donde $p \nmid a'$, $p \nmid b'$, y $a + b = p^e(a' + p^{e'-e} b')$, luego $v_p(a+b) \geq e = \min\{v_p(a), v_p(b)\}$.

   Si además $e < e'$, entonces $p \nmid a' + p^{e'-e} b'$, pues en caso contrario tendríamos que $p \mid a'$, luego $v_p(a+b) = e = \min\{v_p(a), v_p(b)\}$.

Es fácil generalizar los conceptos de máximo común divisor y mínimo común múltiplo:

Definición 2.17 Si A es dominio íntegro, se dice que un elemento d de A es un *máximo común divisor* de unos elementos $a_1, \ldots, a_n$ de A si $d \mid a_1, \ldots, d \mid a_n$ y, para todo c de A que cumpla $c \mid a_1, \ldots, c \mid a_n$, se cumple que $c \mid d$.

Igualmente se dice que m es un *mínimo común múltiplo* de $a_1, \ldots, a_n$ si $a_1 \mid m, \ldots, a_n \mid m$ y, para todo c de A que cumpla $a_1 \mid c, \ldots, a_n \mid c$, se cumple que $m \mid c$.

Notemos que en un dominio íntegro arbitrario no podemos definir el máximo común divisor como el mayor de los divisores comunes porque no tenemos ninguna relación de orden definida en el anillo, pero podemos usar como hemos hecho la definición de divisibilidad, y lo mismo sucede con el concepto de mínimo común múltiplo. Sin embargo, con esta definición no es evidente que existan el máximo común divisor y el mínimo común múltiplo de unos elementos de cualquier dominio íntegro dado. Enseguida veremos que siempre existen en todo dominio de factorización única, pero antes observemos que, como todos los conceptos relacionados con la divisibilidad, estos conceptos están definidos de modo que no distinguen entre elementos asociados. En $\mathbb{N}$, dos números naturales (o cualquier número finito, de hecho) tienen un único máximo común divisor y un único mínimo común múltiplo, pero en un dominio íntegro todo asociado de un máximo común divisor (o mínimo común múltiplo) es otro máximo común

divisor (o mínimo común múltiplo). Esto es inmediato, debido a que elementos asociados tienen los mismos múltiplos y divisores, por lo que uno cumple la definición si y sólo si la cumple el otro. Más aún, si d_1 y d_2 cumplen ambos la definición de máximo común divisor, entonces $d_1 \sim d_2$. Esto se debe a que al aplicar la definición para d_1 con $c = d_2$ se sigue que $d_2 \mid d_1$, e igualmente $d_1 \mid d_2$, luego $d_1 \sim d_2$. Lo mismo sucede con la definición de mínimo común múltiplo.

La prueba de que en un dominio de factorización única siempre existe el máximo común divisor y el mínimo común múltiplo de unos elementos dados es esencialmente la misma que hemos visto para el caso de $\mathbb{N}$:

> *Si $a_1, \dots, a_n$ son elementos no nulos de un dominio de factorización única, entonces tienen máximo común divisor y mínimo común múltiplo. Concretamente:*
>
> *Un máximo común divisor se obtiene multiplicando los divisores primos comunes elevados al menor exponente con que los dividen.*
>
> *Un mínimo común múltiplo se obtiene multiplicando todos los divisores primos (comunes y no comunes) elevados al mayor exponente con que los dividen.*

En efecto, observemos que si $a_1, \dots, a_n$ no son unidades, podemos fijar descomposiciones en factores primos

$$a_i = u_i p_1^{e_{i1}} \cdots p_r^{e_{ir}}$$

de modo que los p_i sean los mismos en todas ellas (no meramente asociados) con los exponentes $e_{ij} \geq 0$ (pero admitimos exponentes nulos porque un mismo primo p_j no tiene por qué aparecer en todos los a_i). Basta tomar

$$d = p_1^{m_1} \cdots p_r^{m_r}, \qquad m = p_1^{M_1} \cdots p_r^{M_r},$$

donde $m_j = \min\{e_{1j}, \dots, e_{nj}\}$, $M_j = \max\{e_{1j}, \dots, e_{nj}\}$.

Es claro entonces que $v_p(d) \leq v_p(a_i)$ para todo i, luego $d \mid a_i$, así como que, si $c \mid a_i$ para todo i, entonces $v_{p_j}(c) \leq v_{p_j}(a_i)$ para todo i, luego se cumple que $v_{p_j}(c) \leq m_j = v_{p_j}(d)$. Esto implica que $v_p(c) \leq v_p(d)$ para todo p, pues lo tenemos probado si p es asociado a un p_j, y en caso contrario $v_p(c) \leq v_p(a_1) = 0$ implica que $v_p(c) = 0$ y la desigualdad se cumple trivialmente. Por consiguiente $c \mid d$ y d es un máximo común divisor de $a_1, \dots, a_r$. Igualmente se prueba que m es un mínimo común múltiplo.

Si algún a_i es una unidad, es claro que un máximo común divisor es 1, mientras que para calcular un mínimo común múltiplo podemos eliminar todos los a_i que sean unidades (y si todas lo son, entonces 1 es un mínimo común múltiplo).

2.4 Dominios euclídeos

Veamos ahora que la demostración del teorema fundamental de la aritmética puede generalizarse para probar que una familia relativamente amplia de dominios íntegros son dominios de factorización única.

Definición 2.18 Un dominio íntegro A es un *dominio euclídeo* si existe una aplicación[8] $\phi : A \setminus \{0\} \longrightarrow \mathbb{N}$ (llamada *norma euclídea*) que cumple las dos propiedades siguientes:

1. Si a y b son elementos no nulos de A, entonces $\phi(a) \leq \phi(ab)$.
2. Para cada D y d en A, con $d \neq 0$, existen c y r en A tales que $D = dc + r$, de modo que o bien $r = 0$ o bien $\phi(r) < \phi(d)$.

Ya conocemos algunos ejemplos de dominios euclídeos:

- *$\mathbb{Z}$ es un dominio euclídeo tomando como norma $\phi(n) = |n|$.*

  En efecto, obviamente se cumple que $|m| \leq |mn|$ (donde m y n son números enteros no nulos) y, dados enteros D y d, con $d \neq 0$, podemos dividir $|D| = |d|c + r$, donde $0 \leq r < |d|$. En principio, c es un número natural, pero cambiándolo por $-c$ si $d < 0$, tenemos un cociente entero tal que $|D| = dc + r$. Si D es positivo ya tenemos que $D = dc + r$. En caso contrario, cambiamos c por $-c$ y tenemos que $D = dc - r$, donde $|-r| < |d|$.

  Conviene observar que el cociente y el resto no son únicos, pero la definición de dominio euclídeo no exige que lo sean. Por ejemplo:

  $$17 = 3 \cdot 5 + 2, \text{ con } |2| < |3|, \qquad 17 = 3 \cdot 6 - 1, \text{con } |-1| < 3.$$

  No es difícil ver que cuando $r \neq 0$, hay exactamente dos cocientes posibles con dos restos posibles, uno positivo y otro negativo, luego la división es única si exigimos que el resto no sea negativo.

- *Si k es un cuerpo, el anillo de polinomios $k[x]$ es un dominio euclídeo tomando como norma $\phi(p) = \operatorname{grad} p$.*

  En efecto, es claro que si p y q son polinomios no nulos, se cumple la relación $\operatorname{grad} p \leq \operatorname{grad} p + \operatorname{grad} q = \operatorname{grad}(pq)$. Por otro lado, en la sección 1.3 hemos probado que todo par de polinomios D y d con $d \neq 0$ tiene división euclídea (teniendo en cuenta que el coeficiente director de d es una unidad en k), porque en un cuerpo todo elemento no nulo es una unidad.

Vamos a probar que todo dominio euclídeo es un dominio de factorización única. En primer lugar observamos lo siguiente:

[8] Esto quiere decir que ϕ es un criterio que a cada elemento a de A le asigna un número natural $\phi(a)$.

Si en un dominio euclídeo dos elementos no nulos cumplen que $a \mid b$ y no son asociados, entonces $\phi(a) < \phi(b)$.

Pongamos que $b = au$. Por la definición de norma euclídea $\phi(a) \leq \phi(b)$. Supongamos que se da la igualdad $\phi(a) = \phi(b)$. Dividimos

$$a = bc + r, \quad \text{con } r = 0 \text{ o } \phi(r) < \phi(b) = \phi(a).$$

Entonces $r = a(1 - uc)$. Si $uc \neq 1$, entonces $r \neq 0$, luego

$$\phi(a) \leq \phi(a(1 - uc)) = \phi(r) < \phi(a),$$

y tenemos una contradicción, luego $uc = 1$, luego u es una unidad, luego $a \sim b$. ∎

Teorema 2.19 *En un dominio euclídeo, todo elemento no nulo ni unitario se descompone en producto de elementos irreducibles.*

Demostración: Observemos que todo elemento no nulo ni unitario a tiene un factor irreducible. En efecto, basta tomar un divisor p que no sea unitario de norma mínima. Si no fuera irreducible, podríamos descomponerlo como $p = bc$, donde b y c no son unidades, pero por el resultado precedente $\phi(b) < \phi(p)$ y b también es un divisor de a que contradice la minimalidad de p.

Así pues, dado un elemento a no nulo ni unitario, podemos tomar un divisor irreducible p_1, de modo que $a = p_1 a_1$. Por el resultado precedente, $\phi(a) > \phi(a_1)$. Si a_1 no es una unidad, tiene un factor irreducible p_2, de modo que $a = p_1 p_2 a_2$ y $\phi(a_1) > \phi(a_2)$. Como no puede haber una sucesión decreciente de números naturales, tras un número finito de pasos tenemos que llegar a un a_r que sea una unidad, y cambiando p_r por $p_r a_r$, que también es irreducible, llegamos a que $a = p_1 \cdots p_r$ es una descomposición de a en factores irreducibles. ∎

Ahora falta probar que las descomposiciones en factores irreducibles son únicas salvo orden o asociación, lo cual se obtiene probando que los elementos irreducibles son primos, lo cual a su vez se basa en la relación de Bezout:

Teorema 2.20 (Relación de Bezout) *Si a y b son dos elementos (que no sean los dos nulos) de un dominio euclídeo, tienen un máximo común divisor $d = (a, b)$ y existen u, v tales que $d = ua + vb$.*

La prueba consiste en observar que el algoritmo de Euclides es válido en cualquier dominio euclídeo. En primer lugar observamos que $(a, b) = (a, b + ca)$, en el sentido de que uno de los dos miembros existe si y sólo si existe el otro, y en tal caso son iguales. El argumento es el mismo empleado en el caso de $\mathbb{N}$ para demostrar el teorema fundamental de la aritmética: ambos pares de elementos tienen los mismos divisores comunes. En particular, si tenemos dos elementos a, b y dividimos $b = ac + r$, entonces $(a, b) = (a, r)$.

Teniendo esto en cuenta, es claro que el algoritmo de Euclides nos lleva siempre a un máximo común divisor de a y b. La tabla siguiente muestra un

ejemplo en $\mathbb{Q}[x]$, pero es evidente que todos los pasos pueden llevarse a cabo en cualquier dominio euclídeo. Ponemos en primer lugar el polinomio de mayor grado (de mayor norma euclídea) y vamos calculando cocientes y restos. En cada paso el grado (la norma) del polinomio obtenido es estrictamente menor que la del polinomio precedente, luego tras un número finito de pasos se tiene que llegar al polinomio nulo.

c	r	u	v
	$x^{13}-3x^{12}+2x^{11}-6x^{10}+2x^9-3x^8+x^7+x^3-3x^2+x-3$	1	0
x^7	$x^6-3x^5+2x^4-6x^3+2x^2-3x+1$	0	1
x^3+x	x^3-3x^2+x-3	1	$-x^7$
$x-3$	x^2+1	$-x^3-x$	$x^{10}+x^8+1$
	0		

Concluimos que un máximo común divisor de

$$\begin{aligned} p_1 &= x^{13}-3x^{12}+2x^{11}-6x^{10}+2x^9-3x^8+x^7+x^3-3x^2+x-3, \\ p_2 &= x^6-3x^5+2x^4-6x^3+2x^2-3x+1 \end{aligned}$$

es $d=x^2+1$, así como que

$$d=-(x^3+1)p_1+(x^{10}+x^8+1)p_2.$$

Ahora, exactamente los mismos argumentos empleados en la prueba del teorema fundamental de la aritmética nos dan los teoremas siguientes:

Teorema 2.21 *En un dominio euclídeo, los elementos irreducibles son primos.*

Teorema 2.22 *Todo dominio euclídeo es un dominio de factorización única.*

2.5 Divisibilidad en anillos de polinomios

Veamos ahora cómo los resultados de la sección precedente nos permiten concebir la aritmética de los anillos de polinomios en los mismos términos que concebimos la de los números enteros, siempre y cuando tengamos presentes las definiciones generales. Por ejemplo, el concepto de "elementos asociados" (que en $\mathbb{N}$ es trivial) se reduce en $\mathbb{Z}$ al hecho de que dos números son asociados si y sólo si se diferencian a lo sumo en su signo, lo cual es a su vez consecuencia de que las unidades en $\mathbb{Z}$ son ± 1. En un anillo de polinomios la situación es distinta, en virtud del teorema siguiente:

Teorema 2.23 *Si A es un dominio íntegro, las unidades de un anillo de polinomios $A[x_1,\ldots,x_n]$ son las unidades de A.*

Demostración: Teniendo en cuenta que $A[x_1,\ldots,x_n]=A[x_1][x_2]\cdots[x_n]$, basta probar el teorema para $n=1$. Sea, pues, $p(x)$ un polinomio de $A[x]$ que sea una unidad. Esto significa que existe otro polinomio $q(x)$ tal que $p(x)q(x)=1$, pero, tomando grados, tenemos que

$$\operatorname{grad} p(x)+\operatorname{grad} q(x)=\operatorname{grad} 1=0.$$

Como los grados son números naturales, necesariamente

$$\operatorname{grad} p(x) = \operatorname{grad} q(x) = 0,$$

luego $p(x)$ y $q(x)$ son, en realidad, dos elementos a, b de A tales que $ab = 1$, luego $p(x) = a$ es una unidad de A.

Recíprocamente, si a es una unidad de A, entonces $aa^{-1} = 1$, y esto vale igualmente considerando a a y a a^{-1} como polinomios de grado 0, luego a es una unidad de $A[x]$. ∎

Así, por ejemplo, las unidades del anillo $\mathbb{Q}[x]$ son todos los polinomios no nulos[9] de grado 0, ya que $\mathbb{Q}$ es un cuerpo y todos sus elementos no nulos son unidades.

Así, a efectos de divisibilidad, al trabajar en $\mathbb{Q}[x]$ debemos considerar que los polinomios

$$3x - 5 \qquad \text{y} \qquad x - 5/3$$

son "esencialmente la misma cosa", pues se diferencian en el factor unitario $\epsilon = 3$. Al igual que en $\mathbb{Z}$, ante la posibilidad de elegir un asociado cualquiera de un número dado, es útil considerar siempre que es posible el asociado positivo, hay un criterio muy sencillo para seleccionar un asociado de un polinomio para evitar duplicidades:

Definición 2.24 Si A es un dominio, un polinomio de $A[x]$ se dice *mónico* si su coeficiente director es 1.

Así, es evidente que, si k es un cuerpo, todo polinomio no nulo de $k[x]$ es asociado a un único polinomio mónico, pues basta dividirlo entre su coeficiente director para obtener un asociado mónico y, si $p(x)$ y $q(x)$ son dos polinomios mónicos asociados, existe un ϵ en k no nulo tal que $p(x) = \epsilon q(x)$, pero entonces el coeficiente director de $p(x)$ es $\epsilon = 1$, luego $p(x) = q(x)$.

Así pues, en lo tocante a la divisibilidad, la traducción correcta de "ser positivo" en $\mathbb{Z}$ a un anillo $k[x]$ es "ser mónico". Por ejemplo, del mismo modo que dos enteros positivos son asociados si y sólo si son iguales, dos polinomios mónicos en un anillo $k[x]$ son asociados si y sólo si son iguales.

Por ejemplo, ahora podemos afirmar que en $k[x]$ todo polinomio no nulo se descompone de forma única salvo el orden como

$$p(x) = ap_1(x)^{e_1} \cdots p_r(x)^{e_r},$$

donde a es un elemento no nulo de k (es una unidad) y cada $p_i(x)$ es un polinomio mónico irreducible.

[9] Aquí también resiulta útil el convenio de considerar que $\operatorname{grad} 0 = -\infty$, pues así podemos decir que las unidades de $\mathbb{Q}[x]$ son los polinomios de grado 0.

No hay ningún criterio general sencillo para reconocer polinomios irreducibles,[10] pero hay algunos casos sencillos que resultan muy útiles. El más elemental es el siguiente:

Teorema 2.25 *Si k es un cuerpo, los polinomios de grado 1 en $k[x]$ son irreducibles.*

DEMOSTRACIÓN: Si $p(x)$ es un polinomio de grado 1, ciertamente no es nulo ni es una unidad. Si tenemos que $p(x) = q(x)r(x)$, entonces se cumple que $\operatorname{grad} q(x) + \operatorname{grad} r(x) = 1$, luego necesariamente uno de los dos factores tiene grado 0 y el otro grado 1, luego uno es una unidad y el otro un asociado de $p(x)$. Así pues, los únicos divisores de $p(x)$ son las unidades y los asociados, y esto significa que $p(x)$ es irreducible. ∎

Ejemplo Aquí tenemos dos descomposiciones de un mismo polinomio en factores irreducibles:

$$2x^2 - 18 = (2x + 6)(x - 3) = (2x - 6)(x + 3).$$

Superficialmente son distintas, pero esencialmente son la misma, ya que

$$2x + 6 = 2(x + 3) \sim x + 3, \qquad x - 3 \sim 2(x - 3) = 2x - 6.$$

Así, alguien que entienda la unicidad de las descomposiciones en primos en el sentido estricto que se da en $\mathbb{N}$, considerará que el polinomio $2x^2 - 18$ no tiene una única descomposición en factores irreducibles en el anillo $\mathbb{Q}[x]$, pero el concepto de asociación que hemos introducido permite ver esta falta de unicidad como una mera apariencia. Para evitar esta aparente falta de unicidad lo más sencillo es considerar únicamente factores mónicos, dejando aparte el coeficiente director:

$$2x^2 - 18 = 2(x + 3)(x - 3)$$

∎

Es fácil determinar si un polinomio $p(x)$ es divisible entre un factor irreducible de la forma $x - a$. En cualquier caso podemos realizar la división euclídea

$$p(x) = c(x)(x - a) + r,$$

donde el resto r tiene que tener grado menor que 1, es decir, tiene que ser un polinomio constante. Específicamente, al evaluar en a obtenemos que $r = p(a)$. Por lo tanto, $x - a \mid p(x)$ si y sólo si $p(a) = 0$.

Definición 2.26 Si A es un anillo y $p(x)$ es un polinomio con coeficientes en A, se dice que un elemento a de A es una *raíz* de $p(x)$ si y sólo si $p(a) = 0$.

[10] Los anillos de polinomios son dominios de factorización única, y en ellos los conceptos aritméticos de "irreducible" y "primo" son equivalentes. Sin embargo, en el contexto de los polinomios es más habitual usar la palabra "irreducible", mientras que en otros dominios de factorización única se usa preferentemente la palabra "primo".

Acabamos de probar que si k es un cuerpo, un polinomio $x-a$ divide a otro $p(x)$ si y sólo si a es raíz de $p(x)$. Más en general, $p(x)$ tiene raíces distintas $a_1, \dots a_s$ en k si y sólo si la descomposición en factores irreducibles de $p(x)$ es de la forma

$$p(x) = a(x-a_1)^{e_1} \cdots (x-a_s)^{e_s} p_{s+1}(x)^{e_{s+1}} \cdots p_r(x)^{e_r},$$

donde los polinomios $p_i(x)$ son mónicos e irreducibles distintos entre sí y distintos de los $x-a_i$. El exponente $e_i > 0$ se llama *multiplicidad* de la raíz a_i en $p(x)$. Una raíz de un polinomio se dice *simple, doble, triple,* etc. según si su multiplicidad es $1, 2, 3, \dots$

Un poco más en general, si A es un dominio íntegro y k es su cuerpo de cocientes, podemos definir igualmente la multiplicidad de una raíz en A de un polinomio $p(x)$ de $A[x]$ considerando a $p(x)$ como polinomio en $k[x]$.

De aquí deducimos una propiedad fundamental de los polinomios:

Teorema 2.27 *Si A es un dominio íntegro, cada polinomio $p(x)$ de $A[x]$ tiene a lo sumo tantas raíces en A como indica su grado. Más precisamente, la suma de las multiplicidades de las raíces de $p(x)$ no puede exceder su grado.*

DEMOSTRACIÓN: Considerando el cuerpo de cocientes k de A y trabajando en $k[x]$, si la suma de las multiplicidades de las raíces superara el grado de $p(x)$, la factorización

$$p(x) = a(x-a_1)^{e_1} \cdots (x-a_s)^{e_s} p_{s+1}(x)^{e_{s+1}} \cdots p_r(x)^{e_r},$$

nos daría una contradicción. ∎

Así pues, los polinomios más sencillos son aquellos que se descomponen en factores lineales (factores de grado 1), como

$$2x^3 - 17x^2 + 40x - 16 = 2(x-1/2)(x-4)^2.$$

A la hora de buscar posibles raíces de un polinomio, a menudo es útil el resultado siguiente:

Teorema 2.28 *Si A es un dominio de factorización única, k es su cuerpo de cocientes y $p(x)$ es un polinomio mónico con coeficientes en A, entonces todas las raíces de $p(x)$ en k están, de hecho, en A.*

DEMOSTRACIÓN: Pongamos que

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

y sea $c = r/s$ una raíz de $p(x)$ en k, donde r, s son elementos de A. Esto significa que

$$(r/s)^n + a_{n-1}(r/s)^{n-1} + \cdots + a_1(r/s) + a_0 = 0.$$

Multiplicando por s^n queda:

$$r^n = -a_{n-1}sr^{n-1} - \cdots - a_1 s^{n-1} r - a_0 s^n.$$

Que c no esté en A equivale a que $s \nmid r$, lo cual a su vez equivale a que exista un primo p en A tal que $v_p(s) > v_p(r)$. Más explícitamente, llamando $e = v_p(r)$, tenemos que p^e es la mayor potencia de p que divide a r, mientras que $p^{e+1} \mid s$. Ahora bien

$$v_p(-a_i s^{n-i} r^i) = v_p(a_i) + (n-i)v_p(s) + iv_p(r) > ne$$

lo que significa que p^{ne+1} divide a todos los sumandos del miembro derecho de la igualdad precedente, pero entonces p^{ne+1} divide a la suma de todos ellos, luego también $p^{ne+1} \mid r^n$, pero esto es imposible, porque $v_p(r^n) = ne$. ∎

Notemos que este teorema generaliza al resultado que obtuvimos al final de la sección 2.1, según el cual si un número natural tiene una raíz n-sima en $\mathbb{Q}$, de hecho la tiene en $\mathbb{N}$. Esto resulta también de aplicar el teorema anterior al polinomio $x^n - m$, donde m es un entero. Pero el teorema anterior se aplica también a polinomios como

$$p(x) = x^3 - 3x^2 + 5x - 15.$$

Si un número racional c es raíz de este polinomio, ahora sabemos tiene que ser un número entero, y en el capítulo precedente vimos que $p(x)$ se puede dividir entre $x - c$ en $\mathbb{Z}[x]$ (porque $x - c$ es mónico) luego podemos expresar $p(x) = (x - c)q(x)$, donde $q(x)$ está en $\mathbb{Z}[x]$. Es claro entonces que, el término independiente de $p(x)$ es el producto de los términos independientes de los dos factores, es decir, que $c \mid 15$. Esto es claramente cierto en general:

Teorema 2.29 *Si A es un dominio de factorización única, k es su cuerpo de cocientes y $p(x)$ es un polinomio mónico con coeficientes en A y c es una raíz de $p(x)$ en k, entonces c está en A y divide al término independiente de $p(x)$.*

Por lo tanto, las únicas raíces racionales que puede tener el polinomio

$$p(x) = x^3 - 3x^2 + 5x - 15$$

son ± 1, ± 3 o ± 5, luego, con un número finito de comprobaciones, podemos determinar si realmente hay alguna.

La regla de Ruffini Hay una forma más práctica de realizar divisiones euclídeas entre polinomios de la forma $x - a$, que se conoce como *regla de Ruffini*. Consideremos por ejemplo la división:

$$\begin{array}{rrrrl|l}
x^3 & -3x^2 & +5x & -15 & \multicolumn{2}{l}{x \ \ -2} \\
\cline{5-6}
-x^3 & +2x^2 & & & \multicolumn{2}{l}{x^2 - x + 3} \\
\cline{1-3}
 & -x^2 & +5x & & & \\
 & x^2 & -2x & & & \\
\cline{2-3}
 & & 3x & -15 & & \\
 & & -3x & +6 & & \\
\cline{3-4}
 & & & -9 & &
\end{array}
\qquad
\begin{array}{r|rrrr}
 & 1 & -3 & 5 & -15 \\
2 & & 2 & -2 & 6 \\
\hline
 & 1 & -1 & 3 & -9
\end{array}$$

A la izquierda está el algoritmo usual, con la pequeña variante consistente en que hemos cambiado de signo los polinomios que resultan de multiplicar el divisor por cada monomio del cociente, de modo que transformamos en sumas las restas que proporcionan el resto en cada paso.

A la derecha hemos aplicado la regla de Ruffini: en la primera fila ponemos los coeficientes del dividendo (cuidando de no omitir los ceros si los hubiera), en la segunda fila a la izquierda ponemos a, bajamos el primer coeficiente del divisor a la tercera fila y, a partir de ahí, el cálculo consiste en multiplicar por a el número de la tercera fila, poner el resultado en la segunda fila a la derecha y sumar. Así, en la tercera fila quedan los coeficientes del cociente y en último lugar el resto.

Dejamos que el lector se convenza por sí mismo de que son dos formas equivalentes de disponer las mismas cuentas. Por ejemplo, la división:

$$\begin{array}{r|rrr|r} & 1 & -3 & 5 & -15 \\ 3 & & 3 & 0 & 15 \\ \hline & 1 & 0 & 5 & 0 \end{array}$$

muestra que

$$x^3 - 3x^2 + 5x - 15 = (x-3)(x^2+5),$$

con lo que, en particular, vemos que 3 es una raíz del polinomio dado. ∎

Ejemplo Observemos la importancia práctica del teorema 2.27. Por ejemplo, si buscamos las soluciones en $\mathbb{Q}$ de una ecuación como

$$x^5 - 3x^4 + 5x^3 - 15x^2 = 0$$

de antemano sabemos que a lo sumo habrá cinco. Ahora bien, la factorización

$$x^5 - 3x^4 + 5x^3 - 15x^2 = x^2(x^3 - 3x^2 + 5x - 15)$$

muestra que $a_1 = 0$ es una raíz doble del polinomio, luego sólo puede haber tres raíces más. Si obtenemos, como hemos visto, que

$$x^5 - 3x^4 + 5x^3 - 15x^2 = x^2(x-3)(x^2+5)$$

vemos que $a_2 = 3$ es otra raíz y, como obviamente, el polinomio $x^2 + 5$ no tiene raíces en $\mathbb{Q}$, porque una raíz a debería cumplir $a^2 = -5 < 0$, y esto es imposible, podemos concluir que la ecuación dada tiene sólo dos soluciones racionales, $a_1 = 0$ (doble) y $a_2 = 3$ (simple) y sabemos que sería inútil buscar más. ∎

Si un polinomio de grado mayor que 1 tiene una raíz a, entonces no es irreducible, pues tiene el factor irreducible $x - a$. El recíproco no es cierto en general, pero sí que es válido para polinomios de grado 2 o 3:

Teorema 2.30 *Un polinomio de grado 2 o 3 es irreducible en un anillo $k[x]$, donde k es un cuerpo, si y sólo si no tiene raíces en k.*

DEMOSTRACIÓN: Si un polinomio de grado 2 no es irreducible, tiene que descomponerse en producto de dos factores de grado 1, y los dos tendrán una raíz cada uno (tal vez iguales), luego el polinomio de partida también tiene raíces.

Si un polinomio de grado 3 no es irreducible, se descompondrá como producto de dos polinomios de grado estrictamente menor, luego uno de ellos tendrá grado 1 y esto nos da una raíz. ∎

Por ejemplo, el polinomio de grado 4

$$p(x) = (x^2 + 1)(x^2 + 2)$$

es obviamente reducible, pero no tiene raíces en $\mathbb{Q}$, pues una raíz de $p(x)$ debería serlo de uno de los dos factores, pero claramente, ninguno de los dos tiene raíces en $\mathbb{Q}$ (una vez más porque un cuadrado no puede ser negativo).

Hemos probado que si k es un cuerpo $k[x]$ es un dominio euclídeo, pero esto no es cierto si el anillo de coeficientes no es un cuerpo:

Teorema 2.31 *Si A es un dominio íntegro, el anillo $A[x]$ es un dominio euclídeo si y sólo si A es un cuerpo.*

DEMOSTRACIÓN: Supongamos que $A[x]$ es un dominio euclídeo. Sea a un elemento no nulo de A y vamos a ver que tiene inverso. Consideremos el máximo común divisor $d = (a, x)$. Como $d \mid a$, tiene que ser un polinomio de grado 0 y si $dp(x) = x$, necesariamente $p(x)$ tiene que ser de la forma $p(x) = bx$, con $db = 1$. Esto significa que d es una unidad de A y, por consiguiente, otro máximo común divisor de a y x es simplemente $d = 1$.

Por la relación de Bezout, existen polinomios tales que $1 = u(x)a + v(x)x$, y evaluando en 0 queda $u(0)a = 1$, luego a es una unidad de A. ∎

En particular, $\mathbb{Z}[x]$ o $\mathbb{Q}[x, y]$ no son dominios euclídeos. Sin embargo, son igualmente dominios de factorización única.[11]

[11] En general, se cumple [Al 3.28] que si A es un dominio de factorización única, entonces $A[x_1, \ldots, x_n]$ también lo es.

Capítulo III

Congruencias

Un tipo de problemas de los que se ocupa la teoría de números consiste en determinar si determinadas ecuaciones tienen soluciones enteras o racionales. Consideremos, por ejemplo, las ecuaciones siguientes:

$$x^2 - 3y^2 = 5, \qquad x^2 - 97y^2 = 1, \qquad 3x^3 - 7y^3 = 1.$$

Por ejemplo, en el caso de la segunda ecuación es obvio que admite las soluciones $(x, y) = (\pm 1, 0)$, pero aún cabe preguntarse si hay alguna más aparte de las obvias, por ejemplo, podemos plantearnos si existen soluciones estrictamente positivas.

Una primera toma de contacto con un problema de este tipo puede consistir en pedirle a un ordenador que dé valores a x e y para ver qué resultados obtenemos. Por ejemplo, si calculamos $x^2 - 3y^2$ para valores de x e y comprendidos entre 0 y 9, obtenemos los valores recogidos en la tabla siguiente:

	0	1	2	3	4	5	6	7	8	9
0	0	−3	−12	−27	−48	−75	−108	−147	−192	−243
1	1	−2	−11	−26	−47	−74	−107	−146	−191	−242
2	4	1	−8	−23	−44	−71	−104	−143	−188	−239
3	9	6	−3	−18	−39	−66	−99	−138	−183	−234
4	16	13	4	−11	−32	−59	−92	−131	−176	−227
5	25	22	13	−2	−23	−50	−83	−122	−167	−218
6	36	33	24	9	−12	−39	−72	−111	−156	−207
7	49	46	37	22	1	−26	−59	−98	−143	−194
8	64	61	52	37	16	−11	−44	−83	−128	−179
9	81	78	69	54	33	6	−27	−66	−111	−162

Vemos así que si, en el miembro derecho, en lugar de un 5 tuviéramos por ejemplo un 1, o un 4 o un 6, la ecuación tendría soluciones enteras, pero con un 5 no hemos encontrado ninguna. Naturalmente, que no hayamos encontrado ninguna con valores para las variables entre 0 y 9 no significa que no pueda haber soluciones de mayor magnitud.

Sería interesante que el lector programara un ordenador para buscar soluciones de cualquiera de las tres ecuaciones para valores de las variables en un determinado rango (uno que su ordenador pueda explorar en un tiempo razonable). Obviamente, no es necesario que el ordenador muestre una tabla gigantesca, sino que basta con que imprima las soluciones que encuentra, si es que encuentra alguna.

Si el lector recorre, por ejemplo, un rango entre 0 y 10 000 000 para cada variable, no encontrará ninguna solución salvo la trivial $(x, y) = (1, 0)$ en el caso de la segunda ecuación.[1] Lo más aproximado que encontrará será la "casi solución"

$$5\,604^2 - 97 \cdot 569^2 = -1.$$

para la segunda ecuación.

De nuevo debemos insistir en que el hecho de que no encontremos soluciones en cualquier rango prefijado de valores para las variables no nos asegura que no las haya fuera de dicho rango. De hecho, una prueba de que no debemos sacar conjeturas precipitadas en casos como éste nos la proporciona la segunda ecuación, que sí que tiene soluciones no triviales, pero la menor de ellas es

$$62\,809\,633^2 - 97 \cdot 6\,377\,352^2 = 1.$$

No es necesario un ordenador para encontrar dicha solución. De hecho, se trata de una ecuación de Pell (véase la introducción), y en el siglo XVII, Pierre de Fermat sabía resolver ecuaciones como ésa. Sin embargo, el propósito de este capítulo es introducir un concepto y unas técnicas que, entre otras muchas aplicaciones, permiten probar que las otras dos ecuaciones no tienen soluciones enteras, o siquiera racionales.

Vamos a dar en primer lugar un argumento elemental para el caso de la primera ecuación. Supongamos que existieran números enteros (x, y) tales que $x^2 - 3y^2 = 5$. Dividamos $x = 3c + r$, donde $r = 0, 1, 2$. Entonces

$$(3c + r)^2 - 3y^2 = 9c^2 + 6cr + r^2 - 3y^2 = 5,$$

luego

$$5 - r^2 = 9c^2 + 6cr - 3y^2 = 3(3c^2 + 2cr - y^2),$$

y así concluimos que $3 \mid 5 - r^2$. Ahora bien, si damos a r los tres valores que puede tomar, nos encontramos con que tendría que suceder que uno de los números $5 - 0^2 = 5$, $5 - 1^2 = 4$, $5 - 2^2 = 1$ fuera múltiplo de 3, y ninguno lo es, así que no puede existir la solución supuesta.

Ejercicio: Probar que la tercera ecuación tampoco tiene soluciones enteras imitando el razonamiento anterior, pero realizando ahora la división euclídea $x = 7c + r$, con $0 \le r < 6$.

[1] Es claro que en el caso de las dos primeras ecuaciones no necesitamos dar a las variables valores negativos. En cuanto a la tercera, vemos que cualquier solución (x, y) debe cumplir que x e y tengan el mismo signo, y si fueran ambas negativas, entonces $(-x, -y)$ sería una solución de la ecuación $3x^3 - 7y^3 = -1$, luego basta dar valores positivos a las variables y buscar soluciones con $3x^3 - 7y^3 = \pm 1$.

3.1 Anillos de restos

Si reflexionamos sobre el procedimiento que hemos seguido para probar que la ecuación $x^2 - 3y^2 = 5$ no tiene soluciones enteras, vemos que ha consistido en dividir los infinitos valores posibles que en principio podría tomar la variable x (todos los números enteros) en tres clases distintas: la correspondiente a los números de la forma $3c$ (los múltiplos de 3), la correspondiente a los números de la forma $3c+1$ y la de los números de la forma $3c+2$. El cociente c podría tomar en principio infinitos valores, pero los hemos eliminado al concluir que tendría que cumplirse la relación $3 \mid 5 - r^2$, es decir, a una propiedad que tendría que cumplir el resto r de x, y esto nos ha permitido resolver el problema porque, mientras que las posibilidades para x o c son infinitas, sólo hay un número finito de posibilidades para el resto r, lo que nos ha permitido analizarlas y descartarlas una por una.

En este capítulo vamos a ver que esta técnica de reducir un problema a los restos de las cantidades implicadas (dejando así un número finito de posibilidades para analizar) puede realizarse de forma sistemática y evitando buena parte de los cálculos explícitos que hemos tenido que hacer en los ejemplos que hemos considerado, especialmente en el caso de la tercera ecuación.

La idea básica consiste en definir que dos números enteros a y b son *congruentes módulo* m, donde m es un tercer número entero, y se representa por $a \equiv b \pmod{m}$, si ambos tienen dan un mismo resto r al dividirlos[2] entre m, es decir, si $a = mc_1 + r$ y $b = mc_2 + r$, con $0 \le r < m$.

Observemos que $a \equiv b \pmod{m}$ si y sólo si $m \mid a - b$.

En efecto, si a y b son congruentes tenemos que

$$a - b = mc_1 - mc_2 = m(c_1 - c_2),$$

luego $m \mid a - b$ y, recíprocamente, si $m \mid (a - b)$, tenemos que $a - b = mk$ o, equivalentemente $a = mk + b$. Si dividimos $b = mc + r$, con $0 \le r < m$, entonces

$$a = mk + mc + r = m(k + c) + r,$$

luego, por la unicidad de la división euclídea, r es también el resto de la división de a entre m, luego $a \equiv b \pmod{m}$.

Así hemos llegado a una caracterización del concepto de congruencia de números enteros que tiene la ventaja de que tiene sentido en anillos arbitrarios, aunque no sean dominios euclídeos, y por ello vamos a tomarla como definición general:

Definición 3.1 Si a, b, m son elementos de un anillo arbitrario, diremos que a es *congruente* con b *módulo* m, y lo representaremos por $a \equiv b \pmod{m}$, si cumplen $m \mid a - b$.

[2]Recordemos que el resto de la división euclídea en $\mathbb{Z}$ es único si exigimos que $0 \le r < m$.

Resumen 3.1: Propiedades básicas de las congruencias

Reflexiva $a \equiv a \pmod{m}$.

Simétrica Si $a \equiv b \pmod{m}$, entonces $b \equiv a \pmod{m}$.

Transitiva Si $a \equiv b \pmod{m}$ y $b \equiv c \pmod{m}$, entonces $a \equiv c \pmod{m}$.

Compatibilidad Si $a \equiv a' \pmod{m}$ y $b \equiv b' \pmod{m}$, entonces

$$a + b \equiv a + b' \pmod{m}, \qquad ab \equiv a'b' \pmod{m}.$$

Ya hemos visto la interpretación que tiene esta relación en el caso de los números enteros. La congruencia módulo 2 (que es la misma que la congruencia módulo -2) divide a los números enteros en dos *clases de restos:*

$$\begin{array}{cccccccc} \ldots & -4, & -2, & 0, & 2, & 4, & \ldots \\ \ldots & -3, & -1, & 1, & 3, & 5, & \ldots \end{array}$$

la formada por los números pares y la formada por los números impares. Similarmente, la congruencia módulo 3 (o módulo -3) divide a los números enteros en tres clases de restos:

$$\begin{array}{cccccccc} \ldots & -6, & -3, & 0, & 3, & 6, & \ldots \\ \ldots & -5, & -2, & 1, & 4, & 7, & \ldots \\ \ldots & -4, & -1, & 2, & 5, & 8, & \ldots \end{array}$$

la de los números de la forma $3c$, la de los de la forma $3c + 1$ y la de los de la forma $3c + 2$.

En general, la congruencia módulo m divide a los números enteros en $|m|$ clases de restos,[3] cada una de las cuales está formada por los números de la forma $mc + r$, para un mismo resto r tal que $0 \leq r < m$.

El resumen 3.1 recoge las propiedades más elementales de las congruencias en un anillo arbitrario. Todas ellas se demuestran fácilmente. Veamos por ejemplo la demostración de las propiedades de compatibilidad con la suma y el producto: Tenemos que $a = mc + a'$, $b = mc' + b'$, luego

$$a + b = m(c + c') + a' + b', \qquad ab = m^2cc' + mcb' + mc'a' + a'b',$$

luego $(a + b) - (a' + b') = m(c + c')$, $ab - a'b' = m(bcc' + cb' + c'a')$, y esto significa que $a + a' \equiv b + b' \pmod{m}$ y $ab \equiv a'b' \pmod{m}$.

Las tres primeras propiedades nos permiten hablar de clases de restos en anillos arbitrarios:

[3] Observemos que, en cualquier anillo unitario, todos los elementos son congruentes módulo ± 1, así como que cada número sólo es congruente consigo mismo módulo 0, por lo que estas congruencias carecen de interés.

Definición 3.2 Si a y m son elementos de un anillo arbitrario A, la *clase de restos* de a módulo m es el conjunto de todos los elementos x del anillo que cumplen la relación $x \equiv a \pmod{m}$. En principio la representaremos por $[a]_m$, pero, cuando podamos suponer m conocido, escribiremos simplemente $\bar{a}$. Representaremos por A_m o también $A/(m)$ al conjunto de todas las clases de restos de A módulo m.

Así por ejemplo, si consideramos la congruencia módulo 2 en $\mathbb{Z}$, tenemos que

$$\begin{aligned} \bar{0} &= \{\ldots\ -4,\ -2,\ 0,\ 2,\ 4,\ \ldots\} \\ \bar{1} &= \{\ldots\ -3,\ -1,\ 1,\ 3,\ 5,\ \ldots\} \end{aligned}$$

y el hecho de que éstas son las dos únicas clases de restos módulo 2 se expresa en la forma $\mathbb{Z}_2 = \{\bar{0}, \bar{1}\}$. Similarmente, si consideramos la congruencia módulo 7, tenemos que

$$\begin{aligned} \bar{0} &= \{\ldots\ -14,\ -7,\ 0,\ 7,\ 14,\ \ldots\} \\ \bar{1} &= \{\ldots\ -13,\ -6,\ 1,\ 8,\ 15,\ \ldots\} \\ \bar{2} &= \{\ldots\ -12,\ -5,\ 2,\ 9,\ 16,\ \ldots\} \\ \bar{3} &= \{\ldots\ -11,\ -4,\ 3,\ 10,\ 17,\ \ldots\} \\ \bar{4} &= \{\ldots\ -10,\ -3,\ 4,\ 11,\ 18,\ \ldots\} \\ \bar{5} &= \{\ldots\ -9,\ -2,\ 5,\ 12,\ 19,\ \ldots\} \\ \bar{6} &= \{\ldots\ -8,\ -1,\ 6,\ 13,\ 20,\ \ldots\} \end{aligned}$$

y $\mathbb{Z}_7 = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}\}$. Ahora bien, debemos tener en cuenta que ésta es sólo una de entre infinitas formas de representar las clases de restos módulo 7, pues también podríamos escribir: $\mathbb{Z}_7 = \{\overline{14}, \overline{-6}, \bar{9}, \overline{-4}, \overline{-9}, \overline{19}, \bar{6}\}$, ya que, por ejemplo, $\bar{3} = \overline{-4}$ son dos formas equivalentes de nombrar la clase formada por los números enteros cuyo resto al ser divididos entre 7 da 3.

En general, cuando consideramos congruencias módulo m en un anillo arbitrario, tenemos que

$$\bar{a} = \bar{b} \quad \text{si y sólo si} \quad a \equiv b \pmod{m}.$$

En efecto, la igualdad de clases de restos $\bar{a} = \bar{b}$ indica que los elementos congruentes con a son los mismos que los congruentes con b módulo m y, como $a \equiv a \pmod{m}$, también $a \equiv b \pmod{m}$. Recíprocamente, si $a \equiv b \pmod{m}$, entonces la simetría y la transitividad de la congruencia implican que los elementos congruentes con a son los mismos que los congruentes con b, luego $\bar{a} = \bar{b}$.

Así, si $n > 0$, las n clases de restos que forman $\mathbb{Z}_n$ pueden expresarse en la forma

$$\mathbb{Z}_n = \{\bar{0}, \ldots, \overline{n-1}\},$$

pero, un poco más en general, cualquier intervalo de n números consecutivos representa también dichas clases:

$$\mathbb{Z}_n = \{\bar{k}, \overline{k+1}, \ldots, \overline{k+n-1}\}.$$

Por ejemplo, a menudo, en lugar de representar los elementos de $\mathbb{Z}_n$ con los primeros números naturales, resulta útil de tomar representantes positivos y negativos, así:

$$\mathbb{Z}_6 = \{\overline{-2}, \overline{-1}, \bar{0}, \bar{1}, \bar{2}, \bar{3}\}.$$

Y ahora llegamos a un hecho crucial:

Teorema 3.3 *Si A es un anillo y m es un elemento de A, entonces el conjunto A_m de las clases de restos de A módulo m es también un anillo con la suma y el producto dadas por*

$$\bar{a} + \bar{b} = \overline{a+b}, \qquad \bar{a} \cdot \bar{b} = \overline{ab}.$$

Si A es un dominio y m no es una unidad de A, lo mismo le sucede a A_m.

Demostración: El hecho fundamental es que las operaciones están bien definidas. Por ejemplo si consideramos en $\mathbb{Z}$ la congruencia módulo 5, tenemos que $\bar{2} = \bar{7}$ y $\overline{11} = \overline{21}$. Si al calcular la suma de ambas clases como $\bar{2} + \overline{11} = \overline{13}$ y $\bar{7} + \overline{21} = \overline{28}$ pudiéramos llegar a resultados distintos, no podríamos decir cuál de ellos es realmente la suma que pretendemos calcular, pero sucede que $\overline{13} = \overline{28} = \bar{3}$ y, en general, partamos de la representación de las clases que partamos, llegamos siempre a la misma suma, y lo mismo vale para el producto:

$$\bar{2} \cdot \overline{11} = \overline{22}, \qquad \bar{7} \cdot \overline{21} = \overline{147},$$

pero $\overline{22} = \overline{147} = \bar{2}$. Esto es precisamente lo que afirman las propiedades de compatibilidad recogidas en el resumen 3.1: si tenemos dos pares de elementos congruentes entre sí, al sumar o multiplicar un elemento de cada par obtenemos elementos congruentes entre sí, de modo que determinan una misma clase de restos.

Una vez justificado que es posible sumar y multiplicar clases de restos, la comprobación de que estas operaciones cumplen las propiedades de la definición de anillo (o de dominio) es trivial. Veamos por ejemplo la propiedad distributiva:

$$\bar{a}(\bar{b} + \bar{c}) = \bar{a}(\overline{b+c}) = \overline{a(b+c)} = \overline{ab + ac} = \overline{ab} + \overline{ac} = \bar{a}\bar{b} + \bar{a}\bar{c}.$$

Igualmente se comprueban todas las demás. En particular, el elemento neutro para la suma es la clase $0 = \bar{0}$ y el opuesto de un elemento es $-\bar{a} = \overline{-a}$. Si A es unitario, el elemento neutro para el producto es $1 = \bar{1}$ y si un elemento a es una unidad, entonces $\bar{a}^{-1} = \overline{a^{-1}}$.

Observemos también que para que A_m sea un dominio debe cumplirse que $\bar{1} \neq \bar{0}$, lo cual sucede siempre que m no es una unidad de A. ∎

En particular, ahora nos encontramos con una familia de infinitos dominios[4] finitos $\mathbb{Z}_n$, los anillos de restos módulo n, para $n > 1$. Vamos a explorarlos un poco:

[4]Notemos que $\mathbb{Z}_1 = \{\bar{0}\}$ no es un dominio porque no cumple $\bar{1} \neq \bar{0}$ y $\mathbb{Z}_0$ es esencialmente el mismo anillo que $\mathbb{Z}$. Por otra parte, $\mathbb{Z}_n = \mathbb{Z}_{-n}$.

Ejemplos de álgebra modular Según ya hemos visto, el anillo de clases de restos $\mathbb{Z}_2 = \{\bar{0}, \bar{1}\}$ tiene dos elementos, que se suman y se multiplican de acuerdo con las tablas siguientes:

$$\begin{array}{c|cc} + & \bar{0} & \bar{1} \\ \hline \bar{0} & \bar{0} & \bar{1} \\ \bar{1} & \bar{1} & \bar{0} \end{array} \qquad \begin{array}{c|cc} \cdot & \bar{0} & \bar{1} \\ \hline \bar{0} & \bar{0} & \bar{0} \\ \bar{1} & \bar{0} & \bar{1} \end{array}$$

Esta "aritmética módulo 2" tiene una interpretación muy natural. Las ecuaciones

$$\bar{0} + \bar{0} = \bar{1} + \bar{1} = \bar{0}, \qquad \bar{0} + \bar{1} = \bar{1} + \bar{0} = \bar{1}$$

expresan unas propiedades bien conocidas de la suma de números enteros que —con menos tecnicismos— se expresan así:

$$\text{par} + \text{par} = \text{impar} + \text{impar} = \text{par}, \quad \text{par} + \text{impar} = \text{impar} + \text{par} = \text{impar}.$$

Similarmente, el producto en $\mathbb{Z}_2$ expresa las relaciones:

$$\text{par} \times \text{par} = \text{par} \times \text{impar} = \text{impar} \times \text{par} = \text{par}, \quad \text{impar} \times \text{impar} = \text{impar}.$$

Consideremos un caso más "tipico", como pueda ser la aritmética de $\mathbb{Z}_{12}$. Una ecuación como

$$\bar{9} + \bar{5} = \overline{14} = \bar{2}$$

no hace sino expresar algo tan cotidiano como que si un reloj marca las 9, cuando pasen 5 horas marcará las 2. Así, la aritmética módulo 12 no es sino la aritmética usual modificada de forma obvia para que se cumpla la relación $\overline{12} = 0$, es decir, que cada vez que en un cálculo acumulamos 12 unidades éstas "desaparecen" igual que un reloj vuelve a 0 cada vez que acumula 12 horas.

Si queremos distinguir entre las horas anteriores y posteriores al mediodía tendremos que considerar la aritmética de $\mathbb{Z}_{24}$. Una ecuación como $\overline{19} + \bar{8} = \bar{3}$ indica que 8 horas después de las 7 de la tarde son las 3 de la mañana.

Similarmente, la aritmética de $\mathbb{Z}_7$ es la aritmética de los días de la semana, pero no hay razón para que nos restrinjamos a las aritméticas modulares "de uso cotidiano", sino que cualquier número natural n determina su aritmética modular caracterizada por la ecuación $\bar{n} = 0$ (o, más precisamente, por el hecho de que n es el menor número natural no nulo que se identifica con 0).

Conviene no perder de vista que una operación como

$$\bar{3} \cdot \bar{5} = \bar{6} \qquad \text{(módulo 9)}$$

es una operación entre clases de restos, es decir, que lo que afirma es que si tomamos cualquier número cuyo resto módulo 9 sea 3 y lo multiplicamos por cualquier otro número cuyo resto módulo 9 sea 5 obtendremos un número cuyo resto módulo 9 será 6. El número obtenido dependerá de qué números en concreto hemos multiplicado, pero será necesariamente un número de la clase de restos $\bar{6}$. ∎

3.2 Aplicaciones de las congruencias

Ejemplo Consideremos de nuevo la ecuación $x^2 - 3y^2 = 5$. Si tuviera una solución entera, entonces, tomando clases módulo 3 se cumpliría que $\bar{x}^2 = \bar{2}$, pero si calculamos los cuadrados de los elementos de $\mathbb{Z}_3$, vemos que ninguno de ellos es $\bar{2}$:

x	$\bar{0}$	$\bar{1}$	$\bar{2}$
x^2	$\bar{0}$	$\bar{1}$	$\bar{1}$

por lo tanto, la ecuación no puede tener soluciones enteras. Así se ve más claramente cómo la aritmética modular nos ha reducido una ecuación con dos variables que pueden tomar infinitos valores a otra de una variable que sólo puede tomar tres valores y que, por consiguiente, puede resolverse por comprobación directa. Además, todos los cálculos que teníamos que hacer al presentar el argumento sin congruencias han desaparecido.

Más aún vamos a comprobar que la ecuación no tiene soluciones racionales. Una solución racional sería de la forma x/z, y/z, para ciertos números enteros x, y, z, con $z \neq 0$, los cuales cumplirían:

$$x^2 - 3y^2 = 5z^2.$$

Tomando clases módulo 3 la ecuación se convierte en

$$x^2 = 2z^2,$$

pero la tabla siguiente muestra que esta igualdad sólo puede darse si $\bar{x} = \bar{z} = 0$:

x, z	$\bar{0}$	$\bar{1}$	$\bar{2}$
x^2	$\bar{0}$	$\bar{1}$	$\bar{1}$
$2z^2$	$\bar{0}$	$\bar{2}$	$\bar{2}$

esto significa que $x = 3u$ y $z = 3v$, luego la ecuación es

$$9u^2 - 3y^2 = 45v^2,$$

que se simplifica hasta $3u^2 - y^2 = 15v^2$, de donde se sigue que $3 \mid y^2$, luego $3 \mid y$, y así resulta que $3 \mid x$, $3 \mid y$, $3 \mid z$, pero esto nos da una contradicción, ya que al tomar la solución racional de partida $(x/z, y/z)$ podemos exigir que x, y, z no sean todos múltiplos de 3, ya que siempre podemos dividirlos entre la máxima potencia de 3 que los divida a todos para pasar a otra representación de la misma solución (con el mismo denominador) y que cumpla esta propiedad. Sin embargo, hemos probado que necesariamente 3 dividirá a todos ellos. ∎

Ejemplo Consideremos de nuevo la ecuación $3x^3 - 7y^3 = 1$.

Supuesto que admite una solución entera, tomamos clases módulo 7 y se convierte en $\bar{3}\bar{x}^3 = \bar{1}$, pero la tabla siguiente muestra que esto es imposible:

x	$-\bar{3}$	$-\bar{2}$	$-\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{3}x^3$	$\bar{3}$	$-\bar{3}$	$-\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{3}$	$-\bar{3}$

Observemos que al expresar las clases con representantes entre -3 y 3 sólo tenemos que comprobar cuatro casos en lugar de 7, ya que el efecto del cambio de signo es el obvio.[5] ∎

Ejercicio: Comprobar usando la aritmética modular que la ecuación del ejemplo precedente no tiene soluciones racionales.

Ejercicio: Comprobar que toda terna pitagórica contiene un múltiplo de 3, un múltiplo de 4 y un múltiplo de 5.

Ejemplo *La ecuación $x^5 = y^4 + 4$ no tiene soluciones enteras.*

En efecto, basta probar que no tiene soluciones módulo 11, para lo cual basta considerar la tabla siguiente:

x, y	$-\bar{5}$	$-\bar{4}$	$-\bar{3}$	$-\bar{2}$	$-\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
y^2	$\bar{3}$	$\bar{5}$	$-\bar{2}$	$\bar{4}$	$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{4}$	$-\bar{2}$	$\bar{5}$	$\bar{3}$
y^4	$-\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{5}$	$\bar{4}$	$\bar{3}$	$-\bar{2}$
x^5	$-\bar{1}$	$-\bar{1}$	$-\bar{1}$	$\bar{1}$	$-\bar{1}$	$\bar{0}$	$\bar{1}$	$-\bar{1}$	$\bar{1}$	$\bar{1}$	$\bar{1}$
$y^4 + \bar{4}$	$\bar{2}$	$-\bar{4}$	$-\bar{3}$	$-\bar{2}$	$\bar{5}$	$\bar{4}$	$\bar{5}$	$-\bar{2}$	$-\bar{3}$	$-\bar{4}$	$\bar{2}$

Vemos que $x^5 \equiv \pm 1$ (mód 11), mientras que $y^4 + 4 \not\equiv \pm 1$ (mód 11), luego nunca puede darse la igualdad. ∎

El teorema de Gersónides Ahora podemos probar fácilmente el teorema de Gersónides sobre las potencias de 2 y 3 consecutivas:

Teorema 3.4 (Gersónides) *Las únicas potencias de 2 y de 3 consecutivas son $1, 2, 3, 4$ y $8, 9$.*

Demostración: Tenemos que estudiar las soluciones naturales de las ecuaciones $2^x = 3^y \pm 1$. Consideremos en primer lugar el caso $2^x = 3^y + 1$. Las potencias de 3 módulo 8 son:

$$\bar{3}^0 = \bar{1}, \quad \bar{3}^1 = \bar{3}, \quad \bar{3}^2 = \bar{1}, \quad \bar{3}^3 = \bar{3}, \quad \ldots$$

luego $3^y + 1 \equiv 2, 4$ (mód 8). Sin embargo, tenemos también que $2^x \equiv 0$ (mód 8) siempre que $x \geq 3$, luego para que se dé la igualdad tiene que ser $x = 0, 1, 2$, pero $x = 0$ es claramente imposible, y nos quedan las soluciones $2^1 = 3^0 + 1$, $2^2 = 3^1 + 1$.

Consideremos ahora la ecuación $2^x = 3^y - 1$. Ahora $3^y - 1 \equiv 0, 2$ (mód 8). Más concretamente, cuando y es impar se da el caso $3^y - 1 \equiv 2$ (mód 8), luego si y es impar la igualdad exige como antes que $x = 0, 1, 2$ pero en realidad vemos que sólo puede darse el caso $x = 1$, para el cual obtenemos la solución $2^1 = 3^1 - 1$.

[5] Notemos que, a la hora de hacer cálculos, conviene ir reduciendo los resultados parciales a medida que los vamos obteniendo. Por ejemplo, para calcular $\bar{3} \cdot \bar{3}^3$, en lugar de calcular $\overline{81}$ y luego tener que calcular el resto módulo 7, es más práctico calcular

$$\bar{3} \cdot \bar{3} \cdot \bar{3} \cdot \bar{3} = \bar{3} \cdot \bar{3} \cdot \bar{9} = \bar{3} \cdot \bar{3} \cdot \bar{2} = \bar{3} \cdot \bar{6} = \bar{3} \cdot (-\bar{1}) = -\bar{3}.$$

Supongamos ahora que $y = 2k$, de modo que la ecuación se convierte en

$$2^x = 3^{2k} - 1 = (3^k - 1)(3^k + 1).$$

Para que esto suceda es necesario que $3^k + 1$ sea potencia de 2, es decir, que se cumpla $2^r = 3^k + 1$, pero ya conocemos todas las soluciones de este caso, que corresponden a $r = 1, 2$, $k = 0, 1$, luego tiene que ser $y = 0, 2$, si bien $y = 0$ es claramente imposible y sólo queda la solución $2^3 = 3^2 - 1$. ∎

Sumas de cuatro cuadrados Es fácil ver que no todo número natural se puede expresar como suma de dos o de tres cuadrados, pero resulta que siempre se puede expresar como suma de cuatro:

Teorema 3.5 (Lagrange) *Todo número natural es suma de cuatro cuadrados.*

Demostración: En primer lugar observamos que el producto de dos números naturales expresables como suma de cuatro cuadrados es también suma de cuatro cuadrados. Esto es consecuencia de la identidad siguiente:

$$\begin{aligned}&(x^2+y^2+z^2+w^2)(x'^2+y'^2+z'^2+w'^2)\\ &= (xx'-yy'-zz'-ww')^2+(xy'+yx'+zw'-wz')^2\\ &+(xz'+zx'+wy'-yw')^2+(xw'+wx'+yz'-zy')^2\end{aligned} \tag{3.1}$$

Como consecuencia basta demostrar que todo número primo es expresable como suma de cuatro cuadrados.

Por razones técnicas vamos a necesitar la variante que resulta de sustituir x' por $-x'$ en la fórmula anterior:

$$\begin{aligned}&(x^2+y^2+z^2+w^2)(x'^2+y'^2+z'^2+w'^2)\\ &= (-xx'-yy'-zz'-ww')^2+(xy'-yx'+zw'-wz')^2\\ &+(xz'-zx'+wy'-yw')^2+(xw'-wx'+yz'-zy')^2.\end{aligned} \tag{3.2}$$

Como $2 = 1^2 + 1^2 + 0^2 + 0^2$, basta probar que todo primo impar p es suma de cuatro cuadrados.

Los números x^2 con $0 \le x \le \frac{1}{2}(p-1)$ son incongruentes módulo p, e igualmente ocurre con $-1 - y^2$ con $0 \le y \le \frac{1}{2}(p-1)$. Como en total son $p + 1$, existen x, y en estas condiciones tales que

$$x^2 \equiv -1 - y^2 \pmod{p}.$$

Equivalentemente, existe un natural m tal que

$$mp = x^2 + y^2 + 1 < 1 + 2\Big(\frac{1}{2}p\Big)^2 < p^2,$$

luego $0 < m < p$.

Sea r el menor natural no nulo tal que existen números enteros x, y, z, w que cumplan $rp = x^2 + y^2 + z^2 + w^2$. Como m cumple esto, será $r \leq m < p$. Necesariamente r es impar, pues si fuera par, 0, 2 o 4 de los x, y, z, w serían pares y, reordenándolos, podríamos exigir que $x+y$, $x-y$, $z+w$ y $z-w$ fueran pares. Entonces

$$\frac{1}{2}rp = \big(\frac{1}{2}(x+y)\big)^2 + \big(\frac{1}{2}(x-y)\big)^2 + \big(\frac{1}{2}(z+w)\big)^2 + \big(\frac{1}{2}(z-w)\big)^2,$$

en contradicción con la minimalidad de r.

Nuestro objetivo es probar que $r = 1$. Supongamos que $r > 1$ y sean x', y', z', w' los restos módulo r de x, y, z, w entre $-r/2$ y $r/2$ (es posible ya que r es impar). Claramente

$$n = x'^2 + y'^2 + z'^2 + w'^2 \equiv x^2 + y^2 + z^2 + w^2 = rp \equiv 0 \text{ (mód } r),$$

pero $n > 0$, pues en otro caso $x' = y' = z' = w' = 0$, r dividiría x, y, z, w, luego $r^2 \mid x^2 + y^2 + z^2 + w^2 = rp$, de donde $r \mid p$ y en consecuencia $r = 1$, contra lo supuesto. También es claro que $n < 4(\frac{1}{2}r)^2 = r^2$.

Sea $0 < k < r$ tal que $n = kr$. Por la identidad (3.2),

$$krpr = z_1^2 + z_2^2 + z_3^2 + z_4^2$$

para ciertos naturales z_1, z_2, z_3, z_4 y, teniendo en cuenta cómo se obtienen a partir de x, y, z, w, x', y', z', w', es claro que los cuatro son múltiplos de r (por ejemplo $z_1 = -xx' - yy' - zz' - ww' \equiv -x^2 - y^2 - z^2 - w^2 = -rp \equiv 0$ (mód r)).

Así pues $z_i = rt_i$ y por tanto $r^2kp = r^2t_1^2 + r^2t_2^2 + r^2t_3^2 + r^2t_4^2$, con lo que $kp = t_1^2 + t_2^2 + t_3^2 + t_4^2$, en contra de la minimalidad de r. ∎

Ejercicio: Encontrar un número natural que no sea suma de tres cuadrados.

Ecuaciones de Mordell Reciben este nombre las ecuaciones de la forma

$$y^2 = x^3 + k,$$

donde k es un número entero. Mordell demostró que, para cada valor de k, hay a lo sumo un número finito de soluciones enteras, tal vez ninguna. Veamos un ejemplo sencillo de análisis de una ecuación de Mordell:

Ejemplo *Las únicas soluciones enteras de* $y^2 = x^3 + 16$ *son* $(x, y) = (0, \pm 4)$.

$$x^3 = y^2 - 16 = (y+4)(y-4).$$

Supongamos en primer lugar que y es impar. No perdemos generalidad si suponemos que es $y > 0$, y entonces tiene que ser $y > 4$. Se cumple que $(y+4, y-4) = 1$, pues un factor primo común debería dividir a la resta de ambos números, es decir, a 8, luego sería 2, pero entonces $2 \mid y$, en contra de lo supuesto.

Ahora usamos que si el producto de dos números naturales primos entre sí es un cubo, entonces cada factor tiene que ser un cubo. Así pues,

$$y + 4 = u^3, \quad y - 4 = v^3,$$

luego $u^3 - v^3 = 8$. Pero la distancia entre dos cubos es mayor o igual que la distancia entre dos cubos consecutivos: $(k+1)^3 - k^3 = 3k^2 + 3k + 1 > 8$ salvo si $k = 1$, es decir, salvo en el caso de los cubos $1 < 8$, pero no podemos estar en este caso porque nuestros cubos son impares. Así pues, y es par, luego x también es par. Más aún, $x^3 + 16$ es divisible entre 8, luego $4 \mid y$. Haciendo $y = 4y'$ queda $x^3 = 16(y'^2 - 1)$, luego $4 \mid x$. Hacemos también $x = 4x'$ y la ecuación se reduce a

$$4x'^3 = y'^2 - 1.$$

Esto muestra que y' es impar. Digamos que $y' = 2m + 1$, con lo que

$$x'^3 = m^2 + m = m(m+1).$$

Puesto que, obviamente, $(m, m+1) = 1$, ambos factores deben ser cubos.[6]

Ahora bien, los únicos cubos consecutivos son $-1, 0, 1$, luego uno de los dos factores es 0, luego $x' = 0$, luego $x = 0$ y por lo tanto $y = \pm 4$. ∎

Mucho más complicado es el caso $k = 1$:

Ejemplo *Las únicas soluciones enteras de* $y^2 = x^3 + 1$ *son*

$$(x, y) = (-1, 0), \ (0, \pm 1), \ (2, \pm 3).$$

Equivalentemente, el único cubo positivo que precede a un cuadrado es $8 < 9$.

En efecto, escribimos la ecuación en la forma

$$x^3 = y^2 - 1 = (y+1)(y-1).$$

Supongamos en primer lugar que y es par. Entonces $(y+1, y-1) = 1$, pues un primo que dividiera a ambos números dividiría a su diferencia, luego sería 2, y entonces y sería impar. Concluimos entonces que ambos factores son cubos, digamos

$$y + 1 = u^3, \qquad y - 1 = v^3,$$

pero entonces $u^3 - v^3 = 2$, y es fácil ver que los únicos cubos que difieren en dos unidades son $-1 < 1$, luego tiene que ser $v = -1$, $u = 1$, luego $y = 0$ y a su vez $x = -1$.

[6] Aquí es fundamental que -1 es un cubo en $\mathbb{Z}$. Hemos usado en varias ocasiones que si un producto de números naturales primos entre sí es una potencia n-sima, entonces cada factor es una potencia n-sima, pero ahí era fundamental que los factores fueran positivos, ya que $(-4)(-9)$ es un cuadrado y ninguno de los factores lo es. Al considerar números enteros, si tenemos que $xy = z^n$, con $(x, y) = 1$, la factorización única nos da que todos los primos dividen a los factores con exponente múltiplo de n, y esto se traduce en que $x = \pm u^n$, $y = \pm v^n$. Ahora, si n es impar, entonces $-u^n = (-u)^n$ y podemos concluir igualmente que los factores son potencias n-simas.

Supongamos ahora que y es impar. Entonces $(y+1, y-1) = 2$. Por otra parte, uno de los dos números $y \pm 1$ tiene que ser múltiplo de 4. Cambiando y por $-y$ si es preciso, podemos suponer que $y \equiv 1 \pmod 4$. Consecuentemente $y+1 \equiv 2 \pmod 4$, $y-1 \equiv 0 \pmod 4$ y así:

$$\left(\frac{x}{2}\right)^3 = \frac{y+1}{2}\frac{y-1}{4}$$

y los dos factores son primos entre sí, luego ambos son cubos. Pongamos que

$$\frac{y+1}{2} = a^3, \qquad \frac{y-1}{2} = b^3,$$

luego $2a^3 - 1 = y = 4b^3 + 1$. Así pues, y tiene que ser de la forma $4b^3+1$, donde b es (parte de) una solución entera de la ecuación

$$a^3 - 2b^3 = 1.$$

Queremos probar que y sólo puede tomar los valores ± 1, ± 3, pero recordemos que hemos elegido el signo de y para que sea $y \equiv 1 \pmod 4$, por lo que en realidad hay que probar que sólo puede tomar los valores $1, -3$. Para ello basta probar que las únicas soluciones de esta última ecuación son $(a, b) = (1, 0), (-1, -1)$. Esto es un caso particular del ejemplo siguiente. ∎

Ejemplo *La ecuación $x^3 + y^3 = 2z^3$ no tiene soluciones enteras con $x \neq \pm y$.*

Notemos que en particular, las únicas soluciones de la ecuación $x^3 - 1 = 2z^3$ (es decir, haciendo $y = -1$), son las dadas por $x = \pm 1$, es decir, $(x, z) = (1, 0)$, $(-1, -1)$, tal y como necesitamos para completar el ejemplo precedente.

Llamamos soluciones triviales de la ecuación a las que cumplen $x = \pm y$. Si un primo divide a x, y, también divide a z, y podemos simplificarlo para obtener otra solución que será trivial si y sólo si lo era la dada, luego si existiera una solución no trivial, habría una con $(x, y) = 1$, así que no perdemos generalidad si suponemos que es así. Además tienen que ser impares, porque si uno fuera par también lo sería el otro. Por la simetría de x e y en la ecuación podemos suponer también que $y < x$. Llamemos

$$u = \frac{x+y}{2}, \qquad v = \frac{x-y}{2}.$$

Son enteros primos entre sí (un divisor primo común dividiría a su suma y a su diferencia, es decir, a x e y) y, como partimos de una solución no trivial, $v > 0$. En términos de u y v, la ecuación equivale a

$$u(u^2 + 3v^2) = z^3.$$

(Sustituyendo u y v por sus definiciones se comprueba que el miembro izquierdo es $(x^3+y^3)/2$.) Basta probar que no existen soluciones enteras de esta ecuación con $v > 0$.

Supongamos en primer lugar que $3 \nmid u$. Entonces $(u, u^2+3v^2) = 1$, luego los dos factores del miembro izquierdo de la ecuación deben ser cubos. Pongamos que $u = m^3$ y $u^2 + 3v^2 = n^3$, con $(m, n) = 1$. Entonces, llamando $t = n - m^2$,

$$3v^2 = n^3 - m^6 = (t + m^2)^3 - m^6 = t^3 + 3t^2m^2 + 3tm^4,$$

luego

$$t(t^2 + 3tm^2 + 3m^4) = 3v^2.$$

Esto implica que 3 divide a uno de los dos factores, pero es claro que de hecho los divide a ambos, luego $3 \mid t$, $9 \mid 3v^2$, luego $3 \mid v$, pero $3 \nmid m$, o de lo contrario $3 \mid (m, n) = 1$, luego $9 \nmid t^2 + 3tm^2 + 3m^4$, luego $9 \mid t$. Pongamos que $t = 9e$, $v = 3f > 0$, con lo que

$$9e(81e^2 + 273m^2 + 3m^4) = 28f^2,$$

luego

$$e(27e^2 + 9em^2 + m^4) = f^2.$$

Los dos factores del miembro izquierdo son primos entre sí, pues un divisor primo común dividiría a m y a t, luego a $(m, n) = 1$. Notemos que $e > 0$, pues $t > 0$, porque $m^6 = u^2 < u^2 + 3v^2 = n^3$, luego $m^2 < n$ y $t = n - m^2 > 0$. Por lo tanto

$$e = y^2, \qquad 27e^2 + 9em^2 + m^4 = z^2$$

(donde estos nuevos y, z no son necesariamente los números de los que hemos partido). Llamando $x = m$ obtenemos una ecuación

$$x^4 + 9x^2y^2 + 27y^4 = z^2$$

donde $y > 0$. Basta probar que su única solución entera es $(x, y, z) = (0, 0, 0)$. Nos ocuparemos de ello en el ejemplo siguiente. Ahora vamos a ver que en el caso en que $3 \mid u$ llegamos a la misma ecuación.

Recordemos que tenemos la ecuación $u(u^2+3v^2) = z^3$ con $(u, v) = 1$ y $v > 0$, y ahora suponemos que $3 \mid u$. Entonces $3 \mid z$, pero $9 \nmid u^2 + 3v^2$, luego $9 \mid u$. Pongamos que $u = 9u'$ y $z = 3z'$, con lo que

$$u'(27u'^2 + v^2) = z'^3, \qquad (u', 27u'^2 + v^2) = 1.$$

Una vez más concluimos que los dos factores son cubos, digamos

$$u' = m^3, \qquad 27u'^2 + v^2 = n^3,$$

con $(m, n) = 1$. Ahora llamamos $e = n - 3m^2 > 0$, pues

$$27m^6 = 27u'^2 < 27u'^2 + v^2 = n^3,$$

luego $3m^2 < n$. Operando:

$$v^2 = n^3 - 27m^6 = (e + 3m^2)^3 - 27m^6 = e^3 + 9e^2m^2 + 27em^4$$

y así tenemos la ecuación

$$e(e^2 + 9em^2 + 27m^4) = v^2.$$

Como en el caso anterior, tiene que ser $e = x^2$, $e^2 + 9em^2 + 27m^4 = z^2$ y, llamando $y = m$, queda

$$x^4 + 9x^2y^2 + 27y^4 = z^2,$$

con $x > 0$, que es la misma ecuación a la que habíamos llegado, y de nuevo vemos que la existencia de una solución no trivial de la ecuación de partida implica la existencia de una solución no idénticamente nula de esta ecuación. El ejemplo siguiente prueba que no existen tales soluciones. ∎

Ejemplo *La ecuación $x^4 + 9x^2y^2 + 27y^4 = z^2$ no tiene soluciones enteras distintas de $(x, y, z) = (0, 0, 0)$.*

Basta probar que cualquier solución entera cumple $z = 0$, pues esto ya implica que $x = y = 0$. En caso contrario, no perdemos generalidad si suponemos que $z > 0$. Razonaremos por *descenso infinito* (el mismo argumento empleado en la prueba del teorema 2.3), es decir probaremos que si existe una solución con $z > 0$, existe otra con $0 < z' < z$, lo cual implica que no puede existir una solución con $z > 0$ mínimo, y esto equivale a que no existe ninguna.

Si existe una solución no trivial y un primo divide a x y a y, entonces también divide a z, y podemos simplificarlo hasta obtener otra solución no trivial con $0 < z' < z$. Aplicando este proceso un número finito de veces llegamos a una solución con $(x, y) = 1$.

Si $2 \mid x$, entonces y es impar, luego $y^2 \equiv 1 \pmod 4$, luego la ecuación nos da que $z^2 \equiv -1 \pmod 4$ y esta congruencia es imposible. Así pues, x es impar.

Si y es impar, la ecuación implica que z es impar, pero, teniendo en cuenta que todos los cuadrados impares son 1 módulo 8, la ecuación nos da la congruencia $5 \equiv 1 \pmod 8$, lo cual es absurdo. Por lo tanto y es par, y esto hace que z sea impar.

Si $3 \mid x$, la ecuación nos da que $27y^4 \equiv z^2 \pmod{81}$, luego $27 \mid z^2$, luego $9 \mid z$, luego $81 \mid z^2$, luego $3 \mid y$, cuando hemos tomado $(x, y) = 1$. Por lo tanto concluimos que $3 \nmid x$.

Llamemos $y = 2y_0$, con lo que la ecuación es

$$4^2 \cdot 27y_0^4 = z^2 - x^4 - 4 \cdot 9x^2y_0^2,$$

o también

$$4 \cdot 27y_0^4 = \frac{z^2 - x^4}{4} - 9x^2y_0^2.$$

La igualdad siguiente se prueba sin más que operar su miembro izquierdo y aplicar la igualdad anterior:

$$\left(\frac{z + x^2}{2} + 9y_0^2\right)\left(\frac{z - x^2}{2} - 9y_0^2\right) = 27y_0^4.$$

El primer factor es positivo y el producto también, luego el segundo factor también es positivo. Además ambos factores son primos entre sí, pues si un primo p los divide a ambos, también dividirá a su suma y a su diferencia, es decir, $p \mid z$ y $p \mid x^2 + 18y_0^2$. Si $p \mid y_0$ entonces $p \mid x$ y tenemos una contradicción. En caso contrario, como $p \mid 27y_0^2$, tiene que ser $p = 3$, pero entonces $p \mid x$, cuando hemos visto que no es así.

En esta situación es claro que uno de los dos factores será de la forma $27A^4$ y el otro de la forma B^4. Supongamos en primer lugar que

$$\frac{z+x^2}{2} + 9y_0^2 = 27A^4, \qquad \frac{z-x^2}{2} - 9y_0^2 = B^4.$$

Entonces

$$0 \equiv 27A^4 - 18y_0^2 = \frac{z+x^2}{2} - 9y_0^2 = B^4 = x^2 + B^4 \equiv 1 + B^4 \pmod 3,$$

donde hemos usado que $3 \mid x$, luego $x^2 \equiv 1 \pmod 3$, y así llegamos a que $B^4 \equiv -1 \pmod 3$, lo cual es imposible. Por lo tanto, lo que sucede es que

$$\frac{z+x^2}{2} + 9y_0^2 = A^4, \qquad \frac{z-x^2}{2} - 9y_0^2 = 27B^4,$$

con $AB = y_0$. Operando como antes, ahora llegamos a que

$$A^4 - 18y_0^2 = x^2 + 27B^4.$$

Si $2 \mid A$, entonces $2 \nmid B$, y la ecuación nos da $-2y_0^2 \equiv 4 \pmod 8$, pero es fácil ver que esta congruencia es imposible. Por lo tanto, A es impar, y esto implica que B es par.

La ecuación siguiente se comprueba operando su miembro izquierdo:

$$\left(\frac{A^2+x}{2} - \frac{9}{2}B^2\right)\left(\frac{A^2-x}{2} - \frac{9}{2}B^2\right) = 27B^4.$$

El primer factor es positivo, luego el segundo también. En efecto, basta probar que $9B^2 < A^2$, que equivale a que $3B < A$ y, como $AB = y_0$, esto equivale a que $3y_0 < A^2$, o a que $9y_0^2 < A^4$, lo cual es inmediato a partir de la definición de A. Una vez más, de aquí deducimos que

$$\frac{A^2+x}{2} - \frac{9}{2}B^2 = \epsilon C^4, \qquad \frac{A^2-x}{2} - \frac{9}{2}B^2 = \delta D^4,$$

donde ϵ y δ toman uno el valor 1 y otro 27, mientras que $CD = B$. Despejando $x/2$ en ambas ecuaciones e igualando queda: $\epsilon C^4 + \delta D^4 + 9B^2 = A^2$, o también

$$\epsilon C^4 + \delta D^4 + 9C^2D^2 = A^2$$

Llamando $x' = C$ e $y' = D$ o al revés según si $\epsilon = 1$ o $\epsilon = 27$ obtenemos una ecuación

$$x'^4 + 9x'^2y'^2 + 27y'^4 = z'^2,$$

donde $z' = A \le y_0 < y < z$ (donde usamos de nuevo que $y_0 = AB$). Esto completa la prueba del descenso infinito. ∎

El lema LTE Veamos un resultado[7] que permite resolver fácilmente problemas como éstos:

1. ¿Cuál es la mayor potencia de 3 que divide a $5^{18} - 2^{18}$?
2. ¿Cuál es el menor exponente n que cumple que $125 \mid 2^n + 3^n$?
3. ¿Cuál es el exponente de 3 en el número formado por 405 cifras iguales a 3?

Teorema 3.6 *Sean x, y números enteros y p un primo tal que $p \nmid xy$, $p \mid x-y$. Sea $n \geq 1$ un número natural.*

1. *Si p es impar o bien $p = 2$ y $4 \mid x - y$, entonces*

$$v_p(x^n - y^n) = v_p(x - y) + v_p(n).$$

2. *Si $p = 2$ y n es par, $v_2(x^n - y^n) = v_2(x - y) + v_2(x + y) + v_2(n) - 1$.*

Demostración: Supongamos en primer lugar que $p \nmid n$. Usamos la factorización:

$$x^n - y^n = (x - y)(x^{n-1} + x^{n-2}y + \cdots + xy^{n-1} + y^{n-1}).$$

Teniendo en cuenta que $x \equiv y \pmod{p}$, el segundo factor es congruente con nx^{n-1} módulo p, luego no es divisible entre p, y por lo tanto concluimos que $v_p(x^n - y^n) = v_p(x - y)$.

Supongamos ahora que p es impar y $n = p$. Sea $y = x + kp$. Entonces

$$x^{p-1-t}y^t = x^{p-1-t}(x + kp)^t = x^{p-1-t}(x^t + tx^{t-1}kp + \cdots)$$

$$\equiv x^{p-1-t}(x^t + tkpx^{t-1}) \equiv x^{p-1} + tkpx^{p-2} \pmod{p^2}.$$

Por lo tanto, el segundo factor de la descomposición considerada al principio de la prueba es congruente módulo p^2 con

$$x^{p-1} + (x^{p-1} + kpx^{p-2}) + (x^{p-1} + 2kpx^{p-2}) + \cdots \equiv$$

$$px^{p-1} + (1 + 2 + \cdots + p - 1)kpx^{p-2} \equiv px^{p-1} + \frac{p(p-1)}{2}kpx^{p-2} \equiv px^{p-1} \pmod{p^2},$$

luego es divisible entre p, pero no entre p^2. Esto prueba que

$$v_p(x^p - y^p) = v_p(x - y) + 1.$$

Ahora pongamos que $n = p^a m$, donde $p \nmid m$. Entonces

$$v_p(x^n - y^n) = v_p((x^{p^a})^m - (y^{p^a})^m) = v_p(x^{p^a} - y^{p^a}),$$

[7] En inglés se conoce como el *Lifting The Exponent Lemma* (LTE)

por el caso $p \nmid n$ ya probado. Aquí usamos que si $p \mid x-y$, entonces $p \mid x^k - y^k$, precisamente por la factorización considerada al principio. A su vez:

$$v_p(x^n - y^n) = v_p(x^{p^a} - y^{p^a}) = v_p((x^{p^{a-1}})^p - (y^{p^{a-1}})^p) = v_p(x^{p^{a-1}} - y^{p^{a-1}}) + 1$$

por el caso $n = p$ ya probado. Aplicándolo a veces llegamos a que

$$v_p(x^n - y^n) = v_p(x-y) + a.$$

Supongamos ahora que $p = 2$. Eliminando como antes la parte impar del exponente, podemos suponer que $n = 2^a$. Ahora factorizamos:

$$x^{2^a} - y^{2^a} = (x^{2^{a-1}} + y^{2^{a-1}})(x^{2^{a-1}} - y^{2^{a-1}}) = \cdots$$

$$= (x^{2^{a-1}} + y^{2^{a-1}})(x^{2^{a-2}} + y^{2^{a-2}}) \cdots (x^2 + y^2)(x+y)(x-y).$$

Como x e y son impares, $x^2 \equiv y^2 \equiv 1 \pmod 4$, luego, para todo $k \geq 1$, se cumple que $x^{2^k} \equiv y^{2^k} \equiv 1 \pmod 4$, luego $x^{2^k} + y^{2^k} \equiv 2 \pmod 4$. Así, estos factores son divisibles entre 2 con exponente 1, luego, si $a \geq 1$,

$$v_2(x^{2^a} - y^{2^a}) = a - 1 + v_2(x+y) + v_2(x-y).$$

Si $4 \mid x-y$, entonces $x \equiv y \equiv \pm 1 \pmod 4$, luego $x+y \equiv \pm 2 \equiv 2 \pmod 4$, luego $v_2(x+y) = 1$ y llegamos también a la fórmula del enunciado para este caso. ∎

Nota Si el exponente n es impar, aplicando el teorema a $-y$ en lugar de y se obtiene la fórmula

$$v_p(x^n + y^n) = v_p(x+y) + v_p(n)$$

para primos impares (supuesto que $p \mid x+y$ y $p \nmid xy$). ∎

Ahora es fácil responder a las tres preguntas que hemos formulado antes del teorema:

1. $v_3(5^{18} - 2^{18}) = v_3(3) + v_3(18) = 3$.

2. Si n es par, tenemos que

   $$2^n + 3^n \equiv 2^n + (-2)^n \equiv 2^n + 2^n \equiv 2^{n+1} \pmod 5,$$

   luego $5 \nmid 2^n + 3^n$. Por lo tanto, el n buscado tiene que ser impar. Entonces

   $$v_5(2^n + 3^n) = v_5(5) + v_5(n) = v_5(n) + 1,$$

   luego este exponente será mayor o igual que 3 si y sólo si $v_5(n) \geq 2$, luego la respuesta es $n = 25$.

3. El número es

   $$N = \sum_{k=0}^{404} 3 \cdot 10^k = 3 \cdot \frac{10^{405} - 1}{9},$$

   luego $v_3(N) = -1 + v_3(10^{405} - 1^{405}) = -1 + v_3(9) + v_3(405) = 5$.

Veamos otra aplicación:

Ejemplo *Si $x^n+y^n=p^k$, donde p es un primo impar, $n\geq 3$ es también impar y x, $y\geq 0$, entonces n es potencia de p.*

En efecto, si $p \mid x$, la ecuación implica que $p \mid y$, y entonces x/p, y/p cumplen también la ecuación con el mismo exponente n. Repitiendo este paso podemos suponer que $p \nmid xy$. La factorización de $x^n-(-y)^n$ que hemos usado en la prueba del teorema anterior muestra que $x+y$ es potencia de p y claramente no puede ser $x+y=1$, luego $p \mid x+y$. Por el teorema anterior

$$k = v_p(x^n+y^n) = v_p(x+y) + v_p(n).$$

Por otra parte, si $n=p^r m$, con $(m,n)=1$, también se cumple que

$$v_p(x^{p^r}+y^{p^r}) = v_p(x+y)+v_p(n) = k.$$

Así,

$$p^k \mid x^{p^r}+y^{p^r} \mid x^n+y^n = p^k,$$

lo que implica que $n=p^r$. ∎

En particular vemos que ningún primo puede ser suma de dos cubos (positivos), salvo $2=1^3+1^3$ y tal vez 3.

Ejercicio: Probar que 3 no es suma de dos cubos. AYUDA:

$$x^3+y^3 = (x+y)(x^2-xy+y^2).$$

Ejemplo *La única solución de la ecuación $x^n+y^n=3^k$, con $(x,y)=1$, $n\geq 2$, $x,y\geq 0$ es $2^3+1^3=3^2$.*

Obviamente no puede ser $k=0$. Si $3 \mid x$, también $3 \mid y$, en contradicción con $(x,y)=1$, luego $3 \nmid xy$. Si n es par, entonces $x^n \equiv y^n \equiv 1 \pmod 3$, luego $x^n+y^n \equiv 2 \pmod 3$ y tenemos una contradicción. Por lo tanto n es impar. Por el ejemplo anterior, $n=3^r$ y obviamente $x+y=3^s$. Además,

$$k = v_3(x^n+y^n) = s+r,$$

luego, aplicando $3^{()}$, obtenemos que

$$x^n+y^n = 3^k = 3^{s+r} = (x+y)n.$$

Por simetría podemos suponer que $x>y$. Factorizamos:

$$x^n+y^n = (x-y)(x^{n-1}-x^{n-2}y+x^{n-3}y^2-\cdots) = (x+y)n,$$

luego

$$x^{n-1}-x^{n-2}y+x^{n-3}y^2-\cdots = (x-y)(x^{n-2}+x^{n-4}y^2+\cdots) = n.$$

Así podemos concluir que $x^{n-2} \leq n$. Si fuera $x \geq 4$, tendríamos que

$$n \geq x^{n-2} = (1 + (x-1))^{n-2} = 1 + (n-2)(x-1) + \cdots \geq 1 + 3(n-2),$$

luego $2n \leq 5$, pero $n \geq 3$, luego tiene que ser $x \leq 3$, pero sabemos que $3 \nmid x$, luego $1 \leq y < x \leq 2$, lo que sólo deja la posibilidad $x = 2$, $y = 1$. Además

$$n \geq 2^{n-2} = (1+1)^{n-2} \geq 1 + n - 2 + (n-2)^2 = n^2 - 3n + 3,$$

luego $n^2 - 4n + 3 = (n-2)^2 - 1 \leq 0$, luego $n - 2 \leq 1$, luego $n \leq 3$. Como $n \geq 2$ es impar, necesariamente $n = 3$ y tenemos la solución del enunciado. ∎

Criterios de divisibilidad Una aplicación sencilla de la aritmética modular es la obtención de criterios sencillos que determinan si un número dado es divisible o no entre otro a partir de su expresión decimal.

Por ejemplo, si la última cifra de un número n es c, esto equivale a que $n = 10k + c$, entendiendo que $0 \leq c \leq 9$, naturalmente. Teniendo en cuenta que $10 \equiv 0 \pmod{2, 3, 5}$, al tomar congruencias resulta que

$$n \equiv c \pmod{2, 5, 10},$$

es decir, que para calcular el resto de n módulo $2, 5, 10$ podemos cambiar n por su última cifra. En particular:

> *Un número es múltiplo de* 2 *si y sólo si termina en una cifra par, es múltiplo de* 5 *si y sólo si termina en* 0 *o en* 5 *y es múltiplo de* 10 *si y sólo si termina en* 0.

Por otro lado, $10 \equiv 1 \pmod{3, 9}$. Por lo tanto, si

$$n = c_r \cdot 10^r + c_{r-1} \cdot 10^{r-1} + \cdots + c_1 \cdot 10 + c_0,$$

al tomar congruencias módulo 3, 9 resulta que

$$n \equiv c_r + \cdots + c_0 \pmod{3, 9}.$$

Así pues, para calcular el resto de un número módulo 3 o 9 podemos sustituirlo por la suma de sus cifras, y esto puede repetirse sucesivamente hasta llegar a un número de una única cifra. En particular:

> *Un número es múltiplo de* 3 *o* 9 *si y sólo lo es la suma de sus cifras.*

Por otro lado tenemos que $10 \equiv 2 \pmod{4}$, lo cual no ayuda en mucho, pero $100 \equiv 0 \pmod{4}$, así, si expresamos $n = 100k + m$, donde $0 \leq m < 100$, vemos que

$$n \equiv m \pmod{4},$$

pero m no es sino el número formado por las dos últimas cifras de n, por lo que podemos afirmar que para calcular el resto de n módulo 4 basta considerar el número formado por sus dos últimas cifras. En particular:

Un número es múltiplo de 4 *si y sólo si sus dos últimas cifras forman un múltiplo de* 4.

Para comprobar si un número es múltiplo de 6 lo más práctico es aplicar los criterios que determinan si es múltiplo de 2 y de 3.

Para el 7 tenemos que $10 \equiv 3 \pmod 7$, lo cual no da lugar a ningún criterio sencillo, pero eso no impide que podamos usar congruencias para agilizar el cálculo del resto. Por ejemplo:

$$2\,753 \equiv 2 \cdot 3^3 + 0 \cdot 3^2 + (-2) \cdot 3 + 3 \equiv -1 \cdot 2 + 1 + 3 = -2 + 1 + 3 = 2 \pmod 7.$$

Notemos que hemos agrupado $2 \cdot 3^3 = (2 \cdot 3)(3 \cdot 3)$.

Para el 8 lo más que tenemos es que $1\,000 \equiv 0 \pmod 8$, lo que reduce el cálculo del resto módulo 8 de un número al formado por sus tres últimas cifras, lo cual no es mucho.

Ejercicio: Teniendo en cuenta que $10 \equiv -1 \pmod{11}$, dar un criterio sencillo de divisibilidad entre 11.

Una aplicación clásica de la congruencia $10 \equiv 1 \pmod 9$ es la "prueba del 9" usada para chequear si una operación aritmética ha sido realizada sin error, que consiste en repetirla con los restos módulo 9 de los números implicados. Por ejemplo, para comprobar la división

$$3\,426 = 32 \cdot 107 + 2,$$

podemos reducir los números módulo 9 (sumando sus cifras —o, mejor aún, los restos de sus cifras entre -4 y 5— repetidamente):

$$3\,426 \equiv 3 + 4 + 2 - 3 \equiv 6 \equiv -3 \pmod 9, \quad 32 \equiv 5 \pmod 9, \quad 107 \equiv -1 \pmod 9$$

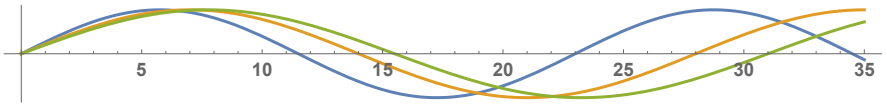
y comprobar que se cumple la congruencia $-3 \equiv 5 \cdot (-1) + 2 \pmod 9$. Si no se cumpliera, indicaría que ha habido un error. Si se cumple, no tenemos la garantía de que no haya error, pero vuelve más improbable esta posibilidad. ∎

Ejercicio: Usar la prueba del 9 para comprobar que esta multiplicación es incorrecta: $325 \cdot 283 = 91\,965$, procurando que los cálculos sean tan simples como sea posible.

3.3 El teorema chino del resto

Biorritmos En muchas ocasiones, el esoterismo y la pseudociencia han planteado problemas sobre números que tienen interés desde un punto de vista puramente matemático, independiente de sus pretendidas interpretaciones adivinatorias, etc. Un ejemplo lo proporciona la teoría sobre los *biorritmos,* una sólida disciplina científica basada en profundos y rigurosos estudios que nadie vio jamás, y que estuvo de moda en la década de 1970.

Según esta teoría, cada ser humano tiene tres indicadores biológicos llamados "biorritmos", uno físico, otro emocional y otro intelectual, que toman el valor 0 en el momento del nacimiento y, a partir de ahí empiezan a oscilar como muestra la figura:



El biorritmo físico realiza un ciclo completo de ascenso, disminución y ascenso de nuevo hasta 0 en un periodo de 23 días —curiosamente, el mismo para cualquier persona—, el emocional requiere 28 días para realizar un ciclo completo y el intelectual 31 días.

A partir de aquí, la teoría de los biorritmos extrae una serie de consecuencias de rigor y características similares a las que se obtienen consultando un horóscopo, pero con aspecto "más matemático". Por ejemplo, si uno tiene que realizar una tarea que requiere un esfuerzo intelectual, será mejor situarla, si es posible, en un momento en el que el biorritmo intelectual tome su valor máximo, es decir, sobre el octavo día de su ciclo, pero si en ese día los otros biorritmos toman valores negativos, eso puede influir negativamente, etc.

Tonterías aparte, la sucesión de los biorritmos plantea problemas numéricos interesantes, unos más fáciles y otros menos. Consideremos en primer lugar el cálculo de los biorritmos de una persona en un día concreto de su vida. Para ello necesitamos conocer su edad expresada en días, entendiendo que en el día de su nacimiento su edad era 0.

Por ejemplo, consideremos el caso de una persona que el 1 de enero de 2020 hubiera cumplido 18 787 días. Entonces, sus biorritmos ese día serían los restos módulo 23, 28 y 31, respectivamente, es decir, $(19, 27, 1)$. Esto significa que el biorritmo físico se encontraría en el día 19 de su periodo de 23 días, el emocional en el día 27 de su ciclo de 28 días y el intelectual en el día 1 de su periodo de 31 días (entendiendo que el primero día de cada periodo es el 0).[8]

Lo primero que podemos preguntarnos es si habrá más días en los que los biorritmos sean $(0, 0, 0)$, cono en el día del nacimiento. Y la respuesta es obviamente positiva. Si el biorritmo físico toma el valor 0 en los múltiplos de 23 días, el emocional en los múltiplos de 28 días y el intelectual en los múltiplos de 31 días, los tres tomarán a la vez el valor 0 en los múltiplos del mínimo común múltiplo de 23, 28 y 31, es decir, cada 19 964 días (aproximadamente cada 54 años y medio).

Una pregunta más sutil es si, cualquier combinación de restos se dará alguna vez. Por ejemplo, la combinación $(5, 7, 7)$ hace que los tres biorritmos se encuentren simultáneamente en su valor máximo. ¿Se da dicha combinación en algún momento?

[8] Los valores de los biorritmos entre -1 y 1 que muestra la gráfica anterior se calcularían como $(\text{sen}(2\pi n/23), \text{sen}(2\pi n/28), \text{sen}(2\pi n/31))$. En nuestro ejemplo, $(-0.89, -0.22, 0.2)$, pero esto es irrelevante para lo que nos interesa.

La respuesta es positiva en el caso de los biorritmos, pero no necesariamente en otros casos similares. Por ejemplo, imaginemos unos "minibiorritmos" que consten únicamente de un ciclo de 6 días y otro de 4 días. Entonces se repetirían cada 12 días, y el ciclo completo sería:

$$(0,0) \to (1,1) \to (2,2) \to (3,3) \to (4,0) \to (5,1) \to$$

$$(0,2) \to (1,3) \to (2,0) \to (3,1) \to (4,2) \to (5,3) \to (0,0)$$

No hace falta calcular explícitamente este ciclo para entender que si sólo pueden darse 12 posibilidades, no pueden darse los 24 casos que en principio serían posibles. Pero viendo la sucesión observamos, por ejemplo, que el caso $(1,2)$ no se da nunca.

El planteamiento general del problema sería el siguiente: dados módulos $m_1, \ldots, m_r$ y números enteros $c_1, \ldots, c_r$, ¿existe un entero x que cumpla las congruencias siguientes?:

$$\begin{aligned} x &\equiv c_1 \ (\text{mód } m_1) \\ &\vdots \qquad\qquad \vdots \\ x &\equiv c_r \ (\text{mód } m_r) \end{aligned}$$

Lo primero que podemos observar a este respecto es que, si m es el mínimo común múltiplo de los módulos m_i y vamos calculando las r-tuplas de restos módulo cada m_i de 0, 1, 2, etc., obtendremos un ciclo de r-tuplas que empezará en $(0, \ldots, 0)$ y volverá a repetirse cuando x tome el valor m, tal y como hemos visto en el caso de los "minibiorritmos".

En efecto, basta observar que $m \equiv 0 \ (\text{mód } m_i)$, luego $km + x \equiv x \ (\text{mód } m_i)$, es decir, que los restos de m son los mismos que los de 0, los de $m+1$ son los mismos que los de 1, los de $m+2$ son los mismos que los de 2, etc.

Pero, más aún, en ese ciclo de longitud m no habrá nunca dos r-tuplas de restos repetidas. Esto se debe a que si x y x' tienen los mismos restos módulo los m_i, esto significa que $m_i \mid x - x'$ para todo i, luego, por definición de mínimo común múltiplo, se cumple que $m \mid x - x'$, luego $x' = x + km$, lo que significa que x y x' no están en el mismo ciclo (salvo que sea $k = 0$ y $x = x'$).

Ahora ya la cuenta es muy simple: si las r-tuplas de restos forman un ciclo de longitud m, de forma que sólo se dan m combinaciones $(c_1, \ldots, c_r)$ de restos, y en principio hay $m_1 \cdots m_r$ combinaciones posibles, resulta que en el ciclo aparecerán todas las combinaciones posibles si y sólo si $m = m_1 \cdots m_r$, pero es fácil ver que el mínimo común múltiplo de unos números es igual a su producto si y sólo si los números son primos entre sí dos a dos.

En realidad, para tener probado el teorema siguiente sólo necesitamos el hecho obvio de que si $m_1, \ldots, m_r$ son primos entre sí dos a dos, entonces su mínimo común múltiplo es su producto:

Teorema 3.7 (Teorema chino del resto) *Sean $m_1, \ldots, m_r$ números enteros primos entre sí dos a dos y sea $m = m_1 \cdots m_r$. Sean $c_1, \ldots, c_r$ números enteros arbitrarios. Entonces el sistema de congruencias*

$$\begin{aligned} x &\equiv c_1 \pmod{m_1} \\ &\vdots \qquad\qquad \vdots \\ x &\equiv c_r \pmod{m_r} \end{aligned}$$

tiene solución y ésta es única módulo m.

Ejemplo El nombre del teorema se debe a que ya era conocido por los chinos en la antigüedad. La referencia más antigua se encuentra en un libro del matemático chino Sun-Zi (maestro Sol), escrito entre los siglos III y V, donde figura este problema:

> *Tenemos cosas en número desconocido. Si las contamos de tres en tres, sobran dos, si las contamos de cinco en cinco, sobran tres y si las contamos de siete en siete, sobran dos. ¿Cuántas cosas hay?*

Esto equivale a resolver las congruencias

$$\begin{aligned} x &\equiv 2 \pmod{3} \\ x &\equiv 3 \pmod{5} \\ x &\equiv 2 \pmod{7} \end{aligned}$$

En realidad la respuesta no es única, a menos que aclaremos que tenemos menos de 105 cosas. El teorema chino del resto nos asegura que el problema tiene solución, pero la prueba que hemos dado no nos proporciona un método para calcularla (aunque siempre puede hacerse "por fuerza bruta", calculando sucesivamente los restos de los números desde 0 hasta a lo sumo 104).

Veamos una técnica para encontrar la solución más directamente. En el ejemplo siguiente mostraremos otra.

1. Calculamos $m_i' = m/m_i$. En nuestro caso $m_1' = 35$, $m_2' = 21$, $m_3' = 15$.
2. Como $(m_i, m_i') = 1$, podemos calcular u_i y v_i tales que $u_i m_i + v_i m_i' = 1$. Para ello podemos usar el algoritmo de Euclides que vimos en el capítulo anterior, si bien en este caso un simple tanteo es suficiente:

$$\begin{aligned} 12 \cdot 3 - 1 \cdot 35 &= 1 \\ -4 \cdot 5 + 1 \cdot 21 &= 1 \\ -2 \cdot 7 + 1 \cdot 15 &= 1 \end{aligned}$$

   Así, $v_i m_i' \equiv 1 \pmod{m_i}$, pero $v_i m_i' \equiv 0 \pmod{m_j}$, para $j \neq i$. Por lo tanto:

3. Una solución es $x = c_1v_1m'_1 + c_2v_2m'_2 + \cdots + c_rv_rm'_r$.

   En nuestro caso,

$$x = 2 \cdot (-1 \cdot 35) + 3 \cdot 1 \cdot 21 + 2 \cdot 1 \cdot 15 = 23.$$

   Como $23 < 105$, se trata de la menor solución posible. La siguiente es 128. ∎

Como los periodos de los biorritmos son primos entre sí dos a dos, el teorema chino del resto nos asegura que cualquier combinación de biorritmos se da en algún momento de la vida de un individuo (al menos si vive un mínimo de 55 años).

Ejercicio: Determinar mediante el algoritmo que acabamos de describir (usando el algoritmo de Euclides) la menor edad en la que los biorritmos de una persona toman los valores $(5, 7, 7)$.

Representaciones de anillos modulares como productos Dado un número natural $m = m_1 \cdots m_r$, donde los m_i son primos entre sí dos a dos, el teorema chino del resto nos dice que podemos representar cada clase de $\mathbb{Z}_m$ como una r-tupla de clases de $\mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_r}$.

Por ejemplo, si queremos operar en $\mathbb{Z}_{60}$, en lugar de representar sus elementos como clases $[n]_{60}$, donde n varía entre 0 y 59, o entre -29 y 30, podemos representar cada clase $[n]_{60}$ como la terna $([n]_4, [n]_3, [n]_5)$.

Por ejemplo,

$$[47]_{60} \leftrightarrow ([3]_4, [2]_3, [2]_5), \qquad [55]_{60} \leftrightarrow ([3]_4, [1]_3, [0]_5),$$

de modo que si tenemos que hacer varios cálculos y no nos importa saber a qué clase corresponden concretamente las coordenadas de los cálculos intermedios, podemos trabajar únicamente con las coordenadas. Por ejemplo:

$$([3]_4, [2]_3, [2]_5) + ([3]_4, [1]_3, [0]_5) = ([2]_4, [0]_3, [2]_5),$$

$$([3]_4, [2]_3, [2]_5) \cdot ([3]_4, [1]_3, [0]_5) = ([1]_4, [2]_3, [0]_5).$$

Aquí estamos usando que si tenemos dos números a y b, los restos módulo 4 o 49 de $a + b$ o ab son, respectivamente las sumas o los productos de los restos correspondientes.

Ejemplo Vamos a calcular $([55]_{60})^4$ usando la expresión anterior. Para calcular una potencia cuarta lo más fácil es elevar dos veces al cuadrado:

$$([3]_4, [1]_3, [0]_5)^2 = ([1]_4, [1]_3, [0]_5), \qquad ([1]_4, [1]_3, [0]_5)^2 = ([1]_4, [1]_3, [0]_5).$$

Para determinar a qué clase módulo 60 corresponde esta terna, vamos a emplear un procedimiento distinto del que hemos explicado anteriormente. Se trata de un método de criba:

1. Partimos de la congruencia con módulo mayor: $x \equiv 0 \pmod 5$, que nos dice que el número que buscamos tiene que ser estar en la sucesión $0, 5, 10, 15, \ldots$

2. Ahora pasamos al módulo siguiente en magnitud $x \equiv 1 \pmod 4$, y vemos que de la sucesión anterior el primer número que cumple esta congruencia es $x = 5$. Por lo tanto, el número que buscamos tiene que ser de la forma $x = 5 + 20k$ (donde $20 = 5 \cdot 4$), luego tiene que estar en la sucesión $5, 25, 45, \ldots$

3. Finalmente, comprobamos que el primer número de la sucesión anterior que cumple $x \equiv 1 \pmod 3$ es $x = 45$, luego $[45]_{60} \leftrightarrow ([1]_4, [1]_3, [0]_5)$. ∎

Ejercicio: Resolver $x \equiv -1 \pmod 8$, $x \equiv 1 \pmod{31}$ por el método de criba que acabamos de describir.

En términos algebraicos, lo que afirma el teorema chino del resto es que la aplicación

$$f : \mathbb{Z}_{60} \longrightarrow \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_5$$

dada por $f([x]_{60}) = ([x]_4, [x]_3, [x]_5)$ es un isomorfismo de anillos, lo cual a su vez significa que f hace corresponder biunívocamente las clases de $\mathbb{Z}_{60}$ con las ternas módulo 4, 3 y 5, respectivamente, y de modo que

$$f(a + b) = f(a) + f(b), \qquad f(ab) = f(a)f(b),$$

entendiendo que las ternas se suman y se multiplican coordenada a coordenada. Naturalmente, esto vale en general para cualquier descomposición de un módulo m en producto de módulos primos entre sí dos a dos.

Pero insistimos en que esto no es sino una forma alternativa de expresar la idea que ya hemos discutido: que trabajar con elementos en un anillo como $\mathbb{Z}_{60}$ es equivalente a trabajar con ternas operándolas componente a componente.

Ejemplo Si sólo nos interesa la edad de una persona para calcular sus biorritmos, es decir, sólo nos interesa el número de días que ha vivido módulo 19 964, en lugar de decir que una persona tiene 18 787 días, podemos decir que tiene $(\bar{19}, \bar{27}, \bar{1})$ días (pues sus biorritmos determinan completamente la edad salvo múltiplos de 19 964). Por ejemplo, si ésa es la edad a 1 de enero de 2020 y queremos saber su edad/biorritmos dos años más tarde, podemos expresar

$$\overline{366 + 365} \leftrightarrow (\overline{18}, \bar{3}, \overline{18})$$

y calcular

$$(\bar{19}, \bar{27}, \bar{1}) + (\overline{18}, \bar{3}, \overline{18}) = (\overline{14}, \bar{2}, \overline{19}).$$

Por la forma en que la hemos obtenido, sabemos que la terna $(\overline{14}, \bar{2}, \overline{19})$ se corresponde con una edad de $18\,787 + 366 + 365 = 19\,518$ días, pero vamos a obtener este resultado por el proceso de criba que hemos descrito en el ejemplo anterior:

1. Partimos de $x \equiv 19 \pmod{31}$, que nos dice que buscamos un número de la forma $19 + 31k$.

2. Vamos a determinar k para que cumpla $19+31k \equiv -9+3k \equiv 2 \pmod{28}$. Esto equivale a $3k \equiv 11 \pmod{28}$. Para despejar k sólo tenemos que recordar la tabla de multiplicar, que nos dice que $3\cdot 9 = 27 \equiv -1 \pmod{28}$, luego $3 \cdot (-9) \equiv 1 \pmod{28}$. Esto significa que $-\bar{9}$ es el inverso de $\bar{3}$ en $\mathbb{Z}_{28}$, luego la ecuación equivale a

$$k \equiv (-9) \cdot 11 \equiv -3 \cdot 33 \equiv -3 \cdot 5 \equiv -15 \equiv 13 \pmod{28}.$$

   Por lo tanto, $x = 19+31\cdot 13 = 412$ cumple las dos primeras congruencias, y buscamos un número de la forma $x = 19 + 31 \cdot 13 + k \cdot 28 \cdot 31$.

3. Ahora buscamos k para que cumpla $19+31\cdot 13+k\cdot 28\cdot 31 \equiv 13 \pmod{23}$ o, equivalentemente $-6k \equiv 6 \pmod{23}$. Así $k \equiv -1 \pmod{23}$, luego nos sirve $k = 22$ y la solución buscada es $x = 19+31\cdot 13+22\cdot 28\cdot 31 = 19\,518$.

■

3.4 El álgebra de los anillos de restos

En los cálculos que ha requerido el ejemplo precedente hemos tenido que resolver la congruencia

$$3k \equiv 11 \pmod{28},$$

que equivale a la ecuación $\bar{3}k = \overline{28}$ en el anillo $\mathbb{Z}_{28}$. Lo hemos conseguido observando que $\bar{3}\cdot(-\bar{9}) = \bar{1}$, es decir, localizando el inverso de $\bar{3}$ en dicho anillo. En general, ante una congruencia de la forma

$$ax \equiv b \pmod{m},$$

no tiene por qué ser cierto que a tenga inverso módulo m y en tal caso la congruencia puede tener solución o no tenerla. Pensemos, por ejemplo, en

$$4x \equiv c \pmod{6}.$$

La tabla siguiente muestra que sólo tiene solución si $c \equiv 0, \pm 2 \pmod{6}$:

x	$-\bar{2}$	$-\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{4}x$	$-\bar{2}$	$\bar{2}$	$\bar{0}$	$-\bar{2}$	$\bar{2}$	$\bar{0}$

Esto implica que 4 no tiene inverso módulo 6 (es decir, que la clase $\bar{4}$ no tiene inverso en $\mathbb{Z}_6$), pues, si lo tuviera, podríamos resolver siempre la ecuación multiplicando ambos miembros por el inverso. La propia tabla muestra por qué $\bar{4}$ no es una unidad de $\mathbb{Z}_6$: porque es un divisor de 0.

En general, un *divisor*[9] *de* 0 en un dominio es un elemento $a \neq 0$ tal que existe otro $b \neq 0$ de modo que $ab = 0$.

[9]Hay que entender que esto significa "divisor de 0 no trivial", pues, de acuerdo con la definición de divisor, todos los elementos de todo anillo dividen a 0.

En estos términos un dominio es un dominio íntegro si y sólo si no tiene divisores de 0. Es inmediato que una unidad no puede ser nunca un divisor de 0, pues si se cumple $ab = 0$ y existe el inverso a^{-1}, multiplicando por a^{-1} obtenemos que, necesariamente, $b = a^{-1}0 = 0$.

Volviendo a nuestro ejemplo, decíamos que el hecho de que $\bar{4} \cdot \bar{3} = 0$ impide que $\bar{4}$ pueda tener un inverso.

Así pues, vemos que los anillos $\mathbb{Z}_m$ no son, en general, dominios íntegros, y así, si queremos operar en ellos con soltura, deberíamos tener claras las respuestas a algunas preguntas básicas:

¿Qué elementos de $\mathbb{Z}_m$ son divisores de 0? ¿Cuáles son unidades?

Si el lector no conoce las respuestas de antemano, sería interesante que considerara algunos ejemplos concretos y formulara algunas conjeturas a partir de ellos:

Ejercicio: Construir las tablas del producto de algunos anillos de restos, como $\mathbb{Z}_6$, $\mathbb{Z}_8$ y $\mathbb{Z}_{10}$. Determinar a partir de ellas cuáles de sus elementos son unidades y cuáles son divisores de 0. Conjeturar un resultado general.

Al lector perezoso le ponemos aquí la tabla de $\mathbb{Z}_6$:

$\cdot$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$
$\bar{2}$	$\bar{0}$	$\bar{2}$	$\bar{4}$	$\bar{0}$	$\bar{2}$	$\bar{4}$
$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{0}$	$\bar{3}$
$\bar{4}$	$\bar{0}$	$\bar{4}$	$\bar{2}$	$\bar{0}$	$\bar{4}$	$\bar{2}$
$\bar{5}$	$\bar{0}$	$\bar{5}$	$\bar{4}$	$\bar{3}$	$\bar{2}$	$\bar{1}$

Empecemos considerando el caso de las unidades. Que un número a sea una unidad módulo m significa que existe un b tal que $ba \equiv 1 \pmod{m}$, lo que a su vez significa que existe un k tal que $ba + km = 1$.

Por la relación de Bezout, una condición suficiente para que esto ocurra es que $(a, m) = 1$, y claramente también es una condición necesaria, pues si $d = (a, m)$ la igualdad $ba + km = 1$ implica que $d \mid 1$, luego $d = 1$.

Por otra parte, la condición $(a, m) = 1$ también es necesaria y suficiente para que a no sea un divisor de 0 módulo m. En efecto, si se cumple sabemos que a es una unidad, luego no puede ser divisor de 0. Recíprocamente, si $d = (a, m) \neq 1$, podemos expresar $a = du$, $m = dv$, con $1 \leq v < m$, luego $\bar{v} \neq \bar{0}$ en $\mathbb{Z}_m$, pero $\bar{v}\bar{a} = \bar{v}\bar{d}\bar{u} = \bar{m}\bar{a} = \bar{0}$, luego $\bar{a}$ es un divisor de 0.

En resumen:

Teorema 3.8 *Si $m > 1$ es un número natural, las unidades de $\mathbb{Z}_m$ son las clases $\bar{a}$ tales que $(a, m) = 1$. Los divisores de 0 de $\mathbb{Z}_m$ son las clases que no son unidades.*

Observemos que no sólo hemos determinado qué elementos de $\mathbb{Z}_m$ son unidades, sino que tenemos un procedimiento para calcular el inverso de cualquier unidad a. Sólo tenemos que encontrar enteros que cumplan $ua + vm = 1$ (por ejemplo, por el algoritmo de Euclides) y entonces $\bar{a}^{-1} = \bar{u}$.

Una consecuencia elemental del teorema anterior:

Teorema 3.9 *Un anillo de restos $\mathbb{Z}_p$ es un cuerpo si y sólo si es un dominio íntegro si y sólo si p es primo.*

Así pues, los anillos $\mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_5$, ... son ejemplos de cuerpos finitos. En ellos podemos sumar, restar, multiplicar y dividir (entre elementos no nulos) con total libertad, como en $\mathbb{Q}$, pero son finitos. Veamos una aplicación:

Teorema 3.10 (Teorema de Wilson) *Si p es primo, entonces*

$$(p-1)! \equiv -1 \text{ (mód } p).$$

DEMOSTRACIÓN: Si a es un elemento no nulo de $\mathbb{Z}_p$, sabemos que tiene un inverso a^{-1}. La clave del argumento es que la igualdad $a = a^{-1}$ equivale a $a^2 = \bar{1}$, o también a que $a^2 - \bar{1} = (a + \bar{1})(a - \bar{1}) = \bar{0}$, luego sólo puede darse cuando $a = \pm\bar{1}$. Así, $\overline{(p-1)!}$ es el producto de todos los elementos no nulos de $\mathbb{Z}_p$, pero sucede que cada uno se cancela con su inverso excepto $\bar{1}$ y $-\bar{1}$, luego el producto total es $-\bar{1}$. ∎

Congruencias lineales Antes de obtener más resultados sobre el álgebra de los anillos de restos, vamos a ver que ya sabemos lo suficiente sobre ellos para determinar cuándo una congruencia de la forma

$$ax \equiv c \text{ (mód } m)$$

tiene solución y cómo encontrarla en caso afirmativo.

Hemos visto que si $(a, m) = 1$, entonces a es una unidad módulo m, luego podemos calcular $\bar{a}^{-1} = \bar{u}$ encontrando los coeficientes que cumplen $ua+vm = 1$ y entonces la solución de la congruencia es $x \equiv uc$ (mód m).

Supongamos ahora que $(a, m) = d > 1$. Una solución de la congruencia cumple $ax = c + km$, de donde se sigue que $d \mid c$. En otras palabras, si $d \nmid c$, podemos concluir que la congruencia no tiene solución. Si $d \mid c$, llamando $a = a'd$, $c = c'd$ y $m = m'd$, vemos que la igualdad $ax = c + km$ es equivalente a $a'x = c' + km'$, es decir, a que

$$a'x \equiv c' \text{ (mód } m'),$$

luego resolver la congruencia inicial equivale a resolver ésta, en la que ya se cumple que $(a', m') = 1$, luego podemos resolverla calculando a'^{-1}. ∎

Ejercicio: Resolver la congruencia $27x \equiv 21$ (mód 69).

Un teorema de Legendre Vamos a probar ahora un teorema de Legendre que requiere usar los resultados principales que hemos probado hasta ahora. Para enunciarlo conviene definir un número *libre de cuadrados* como un número no divisible entre cuadrados (distintos de 1). Equivalentemente, un número es libre de cuadrados si es ± 1 o todos sus divisores primos lo dividen con exponente 1.

Teorema 3.11 (Legendre) *Sean a, b, c números enteros no nulos libres de cuadrados primos entre sí dos a dos y no todos del mismo signo. La ecuación*

$$ax^2 + by^2 + cz^2 = 0$$

tiene soluciones enteras distintas de $x = y = z = 0$ si y sólo si $-ab$ es un cuadrado módulo c, $-ac$ es un cuadrado módulo b y $-bc$ es un cuadrado módulo a.

Demostración: Supongamos que existe una solución no trivial x, y, z. Si dividimos x, y, z entre el máximo común divisor (x, y, z) obtenemos otra solución, así que podemos suponer que x, y, z son primos entre sí. Entonces

$$ax^2 \equiv -by^2 \pmod{c},$$

luego

$$(ax)^2 \equiv -aby^2 \pmod{c}.$$

Ahora bien, tiene que ser $(c, y) = 1$, pues en caso contrario habría un primo p tal que $p \mid c$, $p \mid y$, pero la ecuación implica entonces que $p \mid ax^2$, y por hipótesis $(a, c) = 1$, luego tiene que ser $p \mid x$, y entonces $p^2 \mid ax^2 + by^2 = -cz^2$, pero c es libre de cuadrados, luego sólo contiene a p una vez, luego $p \mid z$, y así $p \mid (x, y, z) = 1$, contradicción.

Por lo tanto, y tiene inverso módulo c, digamos $yy' \equiv 1 \pmod{p}$, con lo que

$$(axy')^2 \equiv -ab \pmod{c},$$

y así tenemos que $-ab$ es un cuadrado módulo c. Por simetría, se cumplen igualmente las otras dos condiciones del enunciado.

Supongamos ahora que se cumplen estas condiciones y vamos a probar que la ecuación tiene una solución no trivial. Para ello probaremos primero que podemos factorizar

$$ax^2 + by^2 + cz^2 \equiv (A_1x + B_1y + C_1z)(A_2x + B_2y + C_2z) \pmod{abc},$$

es decir, que podemos encontrar enteros A_i, B_i, C_i tales que la congruencia anterior es válida para todos los enteros x, y, z. Observemos que el teorema chino del resto implica que basta probarlo módulo a, b y c separadamente. En efecto, si encontramos enteros que cumplan

$$ax^2 + by^2 + cz^2 \equiv (A_1^a x + B_1^a y + C_1^a z)(A_2^a x + B_2^a y + C_2^a z) \pmod{a},$$
$$ax^2 + by^2 + cz^2 \equiv (A_1^b x + B_1^b y + C_1^b z)(A_2^b x + B_2^b y + C_2^b z) \pmod{b},$$
$$ax^2 + by^2 + cz^2 \equiv (A_1^c x + B_1^c y + C_1^c z)(A_2^c x + B_2^c y + C_2^c z) \pmod{c},$$

el teorema chino del resto nos asegura que existen $A_1, A_2, A_3, B_1, B_2, B_3$ tales que

$$A_i \equiv A_i^a \pmod{a}, \qquad A_i \equiv A_i^b \pmod{b}, \qquad A_i \equiv A_i^c \pmod{c},$$

e igualmente con los B_i. Por consiguiente,

$$ax^2 + by^2 + cz^2 \equiv (A_1x + B_1y + C_1z)(A_2x + B_2y + C_2z) \pmod{a},$$

$$ax^2 + by^2 + cz^2 \equiv (A_1x + B_1y + C_1z)(A_2x + B_2y + C_2z) \pmod{b},$$

$$ax^2 + by^2 + cz^2 \equiv (A_1x + B_1y + C_1z)(A_2x + B_2y + C_2z) \pmod{c},$$

y esto implica la congruencia módulo abc. Vamos a considerar el caso módulo a. Por la simetría en las hipótesis, los otros dos casos se cumplirán igualmente.

Por hipótesis existe un entero u tal que $u^2 \equiv -bc \pmod{a}$. Como $(a, b) = 1$, existe un entero b' tal que $bb' \equiv 1 \pmod{a}$, luego

$$ax^2 + by^2 + cz^2 \equiv by^2 + cz^2 \equiv b'(b^2y^2 + bcz^2) \equiv b'((by)^2 - (uz)^2)$$

$$\equiv b'(by + uz)(by - uz) \equiv (b'by + b'uz)(by - uz) \pmod{a},$$

que es la factorización requerida (con $A_1 = 0$).

Como a, b, c no tienen el mismo signo, no perdemos generalidad si suponemos que $a, b > 0$, $c < 0$. Podemos suponer que $|abc| > 1$, pues en caso contrario la ecuación es $x^2 + y^2 - z^2 = 0$, que tiene soluciones no triviales, como $(3, 4, 5)$. Llamemos $\lambda_1 = \sqrt{|bc|}$, $\lambda_2 = \sqrt{|ac|}$, $\lambda_3 = \sqrt{|ab|}$, de modo que $\lambda_1\lambda_2\lambda_3 = |abc|$. Como a, b, c son libres de cuadrados y primos entre sí, no puede ocurrir que todos los λ_i sean enteros (a lo sumo, uno de ellos podría valer 1), luego alguno cumple $E[\lambda_i] < \lambda_i$.

Ahora consideremos todas las ternas (x, y, z) de números naturales que cumplen $0 \le x < \lambda_1$, $0 \le y < \lambda_2$, $0 \le z < \lambda_3$. Si λ_i es entero, hay λ_i posibilidades para la variable i-ésima, pero si no es entero hay $E[\lambda_i] + 1 > \lambda_i$ posibilidades, luego el número de ternas es estrictamente mayor que $\lambda_1\lambda_2\lambda_3 = |abc|$.

Para cada terna, podemos considerar la terna de restos

$$(\overline{A_1x + B_1y + C_1z}, \overline{A_1x + B_1y + C_1z}, \overline{A_1x + B_1y + C_1z}) \in \mathbb{Z}_{|a|} \times \mathbb{Z}_{|b|} \times \mathbb{Z}_{|c|},$$

pero sólo hay $|abc|$ ternas de restos posibles, luego concluimos que tiene que haber dos ternas distintas (x_1, y_1, z_1) y (x_2, y_2, z_2) tales que

$$A_1x_1 + B_1y_1 + C_1z_1 \equiv A_1x_2 + B_1y_2 + C_1z_2 \pmod{abc},$$

luego $x = x_1 - x_2$, $y = y_1 - y_2$, $z = z_1 - z_2$ son enteros no todos nulos que cumplen

$$A_1x + B_1y + C_1z \equiv 0 \pmod{abc}$$

y además $x^2 < |bc|$, $y^2 < |ac|$, $z^2 < ab$. Por la factorización de la ecuación módulo abc tenemos que

$$abc \mid ax^2 + by^2 + cz^2,$$

pero, teniendo en cuenta que $c < 0$ y $a, b > 0$,

$$-|abc| < cz^2 \leq ax^2 + by^2 + cz^2 \leq ax^2 + by^2 < 2|abc|,$$

lo cual sólo deja dos posibilidades: o bien $ax^2 + by^2 + cz^2 = 0$, con lo que ya tenemos la conclusión, o bien $ax^2 + by^2 + cz^2 = |abc| = -abc$. En el segundo caso,

$$ax^2 + by^2 + c(z^2 + ab) = 0.$$

Multiplicando por $z^2 + ab$ queda:

$$(ax^2 + by^2)(z^2 + ab) + c(z^2 + ab)^2 = 0,$$

que equivale a

$$a(xz + by)^2 + b(yz - ax)^2 + c(z^2 + ab)^2 = 0,$$

y la solución es no trivial porque $z^2 + ab > 0$. ∎

Ejercicio: Comprobar que la ecuación $3x^2 + 5y^2 = 7z^2$ no tiene soluciones enteras.

La característica de un anillo Hemos dicho que los cuerpos como $\mathbb{Z}_5$ satisfacen las mismas propiedades algebraicas que $\mathbb{Q}$, por lo que podemos operar en ellos como en $\mathbb{Q}$. Sin embargo, es posible que el lector tenga sus objeciones. En $\mathbb{Z}_5$ pasan cosas como $\bar{3} + \bar{2} = \bar{0}$ que no pasan en $\mathbb{Q}$ o, más aún, en $\mathbb{Q}$ podemos dividir entre 5 y considerar fracciones como $3/5$, pero en $\mathbb{Z}_5$ no podemos, ya que $\bar{5} = \bar{0}$.

La segunda objeción es más aparente que real, pues, tanto en $\mathbb{Q}$ como en $\mathbb{Z}_5$, tenemos la misma restricción a la hora de realizar una división, y es que no podemos dividir entre 0. En cambio, la primera sí que marca una diferencia sustancial entre $\mathbb{Q}$ y $\mathbb{Z}_5$, pero ello no debe provocarnos inseguridad al operar en $\mathbb{Z}_5$, sino que basta comprender exactamente en qué consiste para saber a qué atenernos.

Esencialmente, la peculiaridad de $\mathbb{Z}_5$ es que cumple

$$1 + 1 + 1 + 1 + 1 = 0.$$

Definición 3.12 Se dice que un anillo unitario tiene *característica* $m > 0$ si cumple

$$m \cdot 1 = \overbrace{1 + \cdots + 1}^{m \text{ veces}} = 0$$

y m es el menor número natural para el que se cumple esto. Si esto no sucede para ningún número natural $m > 0$ se dice que el anillo tiene *característica* 0.

En estos términos, $\mathbb{Z}$ o $\mathbb{Q}$ son anillos de característica 0, mientras que $\mathbb{Z}_n$ es un anillo de característica n.

Observemos que un anillo de característica 1 es un anillo que cumple $1 = 0$, y esto implica que el anillo consta únicamente del elemento 0, por lo que podemos dejar de lado este caso. La característica de un dominio será siempre 0 o un número $n \geq 2$.

Ahora podemos probar que los anillos $\mathbb{Z}_n$ no son anillos cualesquiera, sino que están en la base de todos los demás dominios:

Teorema 3.13 *Un dominio A tiene característica 0 si y sólo si contiene a $\mathbb{Z}$ como subdominio, y tiene característica $n \geq 2$ si y sólo si contiene a $\mathbb{Z}_n$ como subdominio.*

DEMOSTRACIÓN: A cada entero n le podemos asignar el elemento $n \cdot 1$ de A, de modo que se cumple

$$(n_1 \cdot 1) + (n_2 \cdot 1) = (n_1 + n_2) \cdot 1 \qquad (n_1 \cdot 1)(n_2 \cdot 1) = (n_1 n_2) \cdot 1,$$

Si A tiene característica 0, entonces siempre tendremos $n \cdot 1 \neq 0$ y, más en general,

$$n_1 \cdot 1 = n_2 \cdot 1 \quad \text{si y sólo si} \quad (n_1 - n_2) \cdot 1 = 0$$
$$\text{si y sólo si} \quad n_1 - n_2 = 0 \quad \text{si y sólo si} \quad n_1 = n_2.$$

Esto significa que podemos identificar cada número entero n con $n \cdot 1$. Al hacerlo la suma y el producto de números enteros se calculan igual usando la aritmética de $\mathbb{Z}$ o la de A, pero así podemos considerar que $\mathbb{Z}$ es un subdominio de A.

Si $\mathbb{Z}$ tiene característica $m \geq 2$ ya no es cierto que $n \cdot 1 \neq 0$. Concretamente, se cumple que $n \cdot 1 = 0$ si y sólo si $m \mid n$.

En efecto, podemos dividir $n = cm + r$, con $0 \leq r < m$, y entonces

$$n \cdot 1 = c \cdot (m \cdot 1) + r \cdot 1 = c \cdot 0 + r \cdot 1 = r \cdot 1.$$

Como m es el menor número natural no nulo tal que $m \cdot 1 = 0$, se cumple que $n \cdot 1 = 0$ si y sólo si $r = 0$, es decir, si y sólo si $m \mid n$.

Un poco más en general, en un anillo A de característica $m > 0$ se cumple que

$$n_1 \cdot 1 = n_2 \cdot 1 \quad \text{si y sólo si} \quad (n_1 - n_2) \cdot 1 = 0 \quad \text{si y sólo si} \quad m \mid n_1 - n_2$$
$$\text{si y sólo si} \quad n_1 \equiv n_2 \pmod{m} \quad \text{si y sólo si} \quad [n_1]_m = [n_2]_m.$$

Y esto se interpreta como que podemos identificar cada clase de restos $[n]_m$ de $\mathbb{Z}_m$ con el elemento $n \cdot 1$ de A, que será el mismo cualquiera que sea el entero n que elijamos en la clase de restos. Igualmente, la suma y el producto de dos clases de $\mathbb{Z}_m$ se calcula igualmente con las operaciones de $\mathbb{Z}_m$ o con las operaciones de A, por lo que en este caso podemos considerar que $\mathbb{Z}_m$ es un subdominio de A. ∎

Como consecuencia:

Teorema 3.14 *Todo dominio íntegro (en particular, todo cuerpo) tiene característica* 0 *o bien característica prima.*

DEMOSTRACIÓN: Si un dominio tiene característica $m > 0$ y no es un número primo, entonces contiene a $\mathbb{Z}_m$, que tiene divisores de 0, luego el dominio dado también los tiene, luego no es íntegro. ∎

Otra consecuencia del teorema 3.13 es que todo cuerpo de característica 0 contiene a $\mathbb{Q}$ como subcuerpo. En efecto, por el teorema 3.13 contiene a $\mathbb{Z}$ y las fracciones en el cuerpo con numerador y denominador enteros forman un subcuerpo que podemos identificar con $\mathbb{Q}$.

En otras palabras, $\mathbb{Q}$ es el menor cuerpo de característica 0, al igual que $\mathbb{Z}_p$ es el menor cuerpo de característica p (para todo primo p).

Así pues, la única "peculiaridad" de un cuerpo como $\mathbb{Z}_5$ es que tiene característica 5, en vez de característica 0, que es la característica de los anillos numéricos $\mathbb{Z}$ y $\mathbb{Q}$. Pero esto no supone ninguna diferencia esencial a la hora de aplicar las reglas usuales de manipulación de expresiones algebraicas que repasamos en el capítulo I. Ninguna de ellas depende de la característica del anillo en que trabajamos.

Entre las peculiaridades que nos podemos encontrar al trabajar con anillos de característica distinta de 0 cabe destacar que en los anillos de característica 2 se cumple la igualdad $2 = 0$, que equivale a $1 = -1$ y a su vez a que $a = -a$, por lo que en estos anillos los signos son redundantes.

Por otra parte, en los anillos de característica prima se cumple una notable propiedad adicional que no es aplicable en los anillos de característica 0. Es una consecuencia del hecho siguiente:

Teorema 3.15 *Si p es un número primo y $0 < m < p$, entonces*

$$\binom{p}{m} \equiv 0 \text{ (mód } p).$$

DEMOSTRACIÓN: Razonamos por inducción sobre m. Si $m = 1$, entonces $\binom{p}{m} = p$, luego efectivamente $p \mid \binom{p}{m}$. Si $p \mid \binom{p}{m}$ y $m + 1 < p$, entonces

$$\binom{p}{m+1} = \frac{p!}{(m+1)!\,(p-m-1)!} = \frac{p-m}{m+1}\,\frac{p!}{m!\,(p-m)!} = \frac{p-m}{m+1}\binom{p}{m}.$$

Así pues, $p \mid (p-m)\binom{p}{m}$, y como $(m+1, p) = 1$, la divisibilidad se conserva al dividir entre $m + 1$, es decir, $p \mid \binom{p}{m+1}$. ∎

De aquí obtenemos lo que algunos llaman "el sueño del aprendiz":

Teorema 3.16 *En un anillo de característica prima p, se cumple que*

$$(a \pm b)^p = a^p \pm b^p.$$

DEMOSTRACIÓN: Basta aplicar la fórmula del binomio de Newton teniendo en cuenta que, por el teorema anterior, todos los números combinatorios son 0 en A excepto el primero y el último. ∎

Un poco más en general, de aquí se sigue inmediatamente que

$$(a \pm b)^{p^n} = a^{p^n} \pm b^{p^n}.$$

Un resultado conocido en el que sí que es relevante la característica del anillo es la fórmula para resolver ecuaciones de segundo grado:

$$ax^2 + bx + c = 0 \quad \text{si y sólo si} \quad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Esta fórmula es válida únicamente en cuerpos de característica distinta de 2, entendida en los términos siguientes:

Teorema 3.17 *Sea k un cuerpo de característica distinta de 2. Entonces una ecuación $ax^2 + bx + c = 0$, con $a \neq 0$, tiene solución en k si y sólo si su discriminante $\Delta = b^2 - 4ac$ tiene raíz cuadrada en k, y en tal caso las soluciones de la ecuación vienen dadas por*

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

DEMOSTRACIÓN: Basta multiplicar la ecuación por $4a$, con lo que resulta ser equivalente a

$$4a^2x^2 + 4abx + 4ac = (2ax + b)^2 - b^2 + 4ac = 0,$$

o también a $(2ax+b)^2 = \Delta$. El polinomio $t^2 - \Delta = 0$ no puede tener más de dos raíces en k y, si una de ellas es $\sqrt{\Delta}$, o bien $\Delta = 0$ y es la única, o bien la otra es $-\sqrt{\Delta}$. En ambos casos podemos decir que sus raíces (si existen) son $\pm\sqrt{\Delta}$.

Así pues, vemos que la ecuación tiene solución si y sólo si existe $\sqrt{\Delta}$ en k, en cuyo caso $2ax - b = \pm\sqrt{\Delta}$, de donde se llega a la fórmula del enunciado (siempre suponiendo que la característica del cuerpo no es 2). ∎

Ejemplo Vamos a resolver la ecuación $x^2 - 4x + 13 = 0$ en el cuerpo $\mathbb{Z}_{41}$.

Su discriminante es $\Delta = 16 - 4 \cdot 13 = -36$. Ahora observamos que

$$-1 \equiv -1 + 41 \cdot 2 \equiv 81 \equiv 9^2 \text{ (mód 41)}$$

luego una raíz cuadrada de -36 es

$$\sqrt{-36} = \sqrt{-1}\sqrt{36} = 9 \cdot 6 \equiv 13 \text{ (mód 41)}.$$

Por lo tanto, las raíces de la ecuación son

$$x = \frac{\bar{4} \pm \overline{13}}{\bar{2}} = \bar{2} \pm \frac{\overline{13}}{\bar{2}}.$$

El inverso de 2 es fácil de calcular:

$$1 \equiv 1 + 41 \equiv 42 = 2 \cdot 21 \text{ (mód 41)}.$$

Así pues: $x \equiv 2 \pm 13 \cdot 21 \equiv -12, 16 \text{ (mód 41)}$. ∎

Ejercicio: Probar que n es un número triangular si y sólo si $8n + 1$ es un cuadrado perfecto.

3.5 Primos de Mersenne

Números perfectos Recordemos la definición de número perfecto que ya presentamos en la introducción. Su origen hay que buscarlo en la numerologia, es decir, de las creencias que atribuyen a los números propiedades mágicas o adivinatorias.

> Un número natural no nulo es *perfecto* si es la suma de sus divisores (naturales) distintos de él mismo.

Por ejemplo, los divisores de 6 son 1, 2 y 3 y se cumple que $6 = 1+2+3$. El número 6 es perfecto. Otro ejemplo de número perfecto es el $28 = 1+2+4+7+14$.

Tal y como señalábamos en la introducción, Euclides recoge en sus *Elementos* el hecho de que si el número

$$1 + 2 + \cdots + 2^n = 2^{n+1} - 1$$

es primo, entonces el número $2^n(2^{n+1} - 1)$ es perfecto.

Para probarlo recordemos la función $\sigma(n)$ que asocia a cada número natural la suma de sus divisores. Como entre ellos incluimos al propio n, tenemos que n es perfecto si y sólo si $\sigma(n) = 2n$. Por lo tanto, como $\sigma(p) = p+1$, en nuestro caso tenemos que $\sigma(2^{n+1} - 1) = 2^{n+1}$, y como σ es multiplicativa,

$$\sigma\big(2^n(2^{n+1} - 1)\big) = \sigma(2^n)\sigma(2^{n+1} - 1) = (2^{n+1} - 1) \cdot 2^{n+1} = 2 \cdot \big(2^n(2^{n+1} - 1)\big).$$

Descartes afirmó (y la primera prueba conocida se debe a Euler) que un número par es perfecto si y sólo si es de la forma indicada por Euclides.

Para verlo tomemos un número par, que lo podremos expresar en la forma $2^n m$, donde $n > 0$ y m es un número impar. Si es perfecto, $\sigma(2^n m) = 2^{n+1}m$. Como σ es multiplicativa

$$(2^{n+1} - 1)\sigma(m) = 2^{n+1}m. \tag{3.3}$$

Como $2^{n+1} - 1$ es impar, el exponente de 2 en $\sigma(m)$ ha de ser $n + 1$, es decir, $\sigma(m) = 2^{n+1}a$ para cierto número natural a. Sustituyendo en (3.3) obtenemos $(2^{n+1} - 1)2^{n+1}a = 2^{n+1}m$, luego

$$(2^{n+1} - 1)a = m. \tag{3.4}$$

Por lo tanto $\sigma(m) = 2^{n+1}a = m + a$. Ahora bien, si $a > 1$ resulta que 1, a y m son divisores distintos de m, luego $\sigma(m)$ debe ser al menos $1 + a + m$, contradicción. Concluimos que $a = 1$ y que $\sigma(m) = m+1$, lo que sólo es posible si m no tiene más divisores que 1 y m, es decir, si m es primo.

Sustituyendo $a = 1$ en (3.4) queda que $m = 2^{n+1} - 1$, luego el número original era $2^n(2^{n+1} - 1)$ con $2^{n+1} - 1$ primo, o sea, de la forma establecida por Euclides. Con un leve cambio de índice podemos enunciar así lo que hemos probado:

> *Un número par es perfecto si y sólo si es de la forma* $2^{n-1}(2^n - 1)$ *y* $2^n - 1$ *es primo.*

Respecto a los números perfectos impares no se conoce ninguno (y se ha comprobado que no hay ninguno menor que $10^{1\,500}$), pero nadie ha demostrado que no existan. Es un problema abierto.

Ahora pasamos a ocuparnos del problema de determinar para qué valores de n se cumple que el número $2^n - 1$ es primo. El lector empírico ya debería haber empezado a recopilar datos. He aquí lo que puede obtenerse sin esforzarse uno mucho:

n	**2**	**3**	4	**5**	6	**7**	8	9	10	¿11?
$2^n - 1$	**3**	**7**	15	**31**	63	**127**	255	511	1023	¿2 047?

La tabla muestra los primeros valores de $2^n - 1$. Es fácil ver que para $n = 2$, 3, 5 y 7 obtenemos primos (el 127 es primo porque no es divisible por primos menores que 12). Tampoco cuesta ver que los números restantes no lo son. Todos son múltiplos de 3 salvo el 511, que es divisible entre 7. El número correspondiente a 11 está entre interrogantes porque ya no es fácil decidir si es primo. Habría que tratar de dividirlo entre todos los primos menores que 45 y hay un total de 14.

Obviamente, es fácil dejar que un ordenador haga los cálculos, pero vamos a ver que, "aguzando el ingenio" todo este asunto puede tratarse en términos que no exceden una capacidad de cálculo moderada asumible sin contar siquiera con una modesta calculadora. Centrémonos de momento en los casos claros.

Vemos que $2^n - 1$ no siempre es primo. ¿Conjetura algo el lector sobre en qué casos lo es? No hay que forzar mucho la imaginación para sospechar que $2^n - 1$ es primo exactamente cuando n lo es.

Tratemos de probar que $2^n - 1$ no es primo si n es compuesto. Para buscar un posible divisor a $2^n - 1$ observemos los divisores que hemos encontrado en nuestros ejemplos:

n	**2**	**3**	2^2	**5**	$2 \cdot 3$	**7**	2^3	3^2	$2 \cdot 5$
$2^n - 1$	**3**	**7**	$3 \cdot 5$	**31**	$3^2 \cdot 7$	**127**	$3 \cdot 5 \cdot 17$	$7 \cdot 73$	$3 \cdot 11 \cdot 31$

Si al lector no le basta esto para conjeturar qué divisores podemos encontrar en el número $2^n - 1$ cuando n no es primo, he aquí una pista más. La factorización siguiente no ha sido hallada por tanteo: $2^{14} - 1 = 16 \cdot 383 = 3 \cdot 43 \cdot 127$. Es fácil encontrar el factor 3, pero ¿cómo se encuentran los factores 43 y 127 en el cociente 5.461?

La mayor pista está en el 127. El 127 es el primo correspondiente a $n = 7$ y 7 es uno de los factores de 14. El otro factor es 2 y su primo correspondiente el 3, que también divide a $2^{14} - 1$. También es claro el caso $n = 10$, cuyos factores son 2 y 5 y en $2^{10} - 1$ aparecen los primos correspondientes, 3 y 31. Vemos que también aparecen otros primos, como el 11 en $2^{10} - 1$ o el 5 y el 17 en $2^8 - 1$, pero eso no importa. La conjetura es, pues:

Si $d \mid n$ *entonces* $2^d - 1 \mid 2^n - 1$.

Aquí tenemos un ejemplo interesante de resultado en el que el uso de congruencias simplifica drásticamente las comprobaciones. En efecto, tomando clases módulo $2^d - 1$ se cumple que $\overline{2}^d = \overline{1}$, luego si $n = dm$

$$\overline{2^n - 1} = \left(\overline{2}^d\right)^m - \overline{1} = \overline{1}^m - \overline{1} = \overline{0},$$

lo que prueba que en efecto $2^d - 1 \mid 2^n - 1$. Con esto ya tenemos que para que $2^n - 1$ sea primo el número n ha de ser primo.

Los números de la forma $M_p = 2^p - 1$ con p primo se llaman *números de Mersenne.*

La cuestión es si todos los números de Mersenne son primos. Antes hemos probado que los primos de Mersenne están en correspondencia con los números pares perfectos. He aquí los números perfectos que hemos encontrado:

p	2	3	5	7
M_p	3	7	31	127
$2^{p-1}M_p$	6	28	496	8 128

Según nuestras conjeturas el número $M_{11} = 2^{11} - 1$ debería ser primo. El lector animoso puede tratar de dividirlo entre los 14 primos menores que 45. Aquí vamos a estudiar el problema pensando más para trabajar menos. Es evidente que $M_{11} = 2\,047$ no es múltiplo de 2, 3 o 5. Vamos a ver si puede ser múltiplo de 7. No sólo vamos a probar que no lo es, sino que vamos a encontrar razones por las que no puede serlo, con lo que podremos descartar muchos más primos aparte del 7. El número $2^{11} - 1$ será múltiplo de 7 si y sólo si al tomar clases módulo 7 se cumple que $\overline{2}^{11} = \overline{1}$. Las potencias de 2 módulo 7 pueden calcularse recurrentemente $\overline{2}^{n+1} = \overline{2}^n\overline{2}$, lo que nos permite reducir las potencias al tiempo que las calculamos: $\overline{2}^2 = \overline{4}$, $\overline{2}^3 = \overline{4} \cdot \overline{2} = \overline{8} = \overline{1}$, $\overline{2}^4 = \overline{1} \cdot \overline{2} = \overline{2}$, etc. Así vamos obteniendo lo siguiente:

n	0	1	2	3	4	5	6	...
$\overline{2}^n$	$\overline{1}$	$\overline{2}$	$\overline{4}$	$\overline{1}$	$\overline{2}$	$\overline{4}$	$\overline{1}$	...

Vemos que el $\overline{1}$ se alcanza cíclicamente en los múltiplos de 3, luego no es necesario llegar hasta el 11: como 11 no es múltiplo de 3, $2^{11} - 1$ no puede ser múltiplo de 7.

Por lo tanto la razón por la que $2^{11} - 1$ no es múltiplo de 7 es que 11 no es múltiplo de 3. ¿De dónde ha salido ese 3? Es obvio que al calcular potencias de

un elemento dado no nulo de $\mathbb{Z}_p$, como es $\bar{2}$, tarde o temprano obtendremos un valor repetido, pues sólo hay un número finito de valores posibles, o sea, dado a en $\mathbb{Z}_p$, existen números naturales $0 < m < n$ tales que $a^m = a^n = a^m \cdot a^{n-m}$, luego $a^{n-m} = 1$. En resumen, ha de haber un natural no nulo d tal que $a^d = 1$. Si tomamos el d menor posible es obvio que el 1 se alcanzará exactamente en los múltiplos de d, que es lo que nos hemos encontrado. En resumen:

> *Dado un primo p y un número a primo con p, existe un número d tal que $p \mid a^n - 1$ si y sólo si $d \mid n$.*

El lector puede calcular los valores de d para $a = 2$ y distintos primos p. No es inmediato, pero tampoco excesivamente laborioso. Como siempre, el lector perezoso tiene a continuación una tabla:

p	3	5	7	11	13	17	19
d	2	4	3	10	12	8	18

¿Encuentra el lector alguna relación entre p y d? Fermat encontró una al estudiar el problema: $d \mid p - 1$. Esto se conoce como *teorema de Fermat.* Para no interrumpir el argumento, posponemos la demostración de este hecho. Aceptándolo, tenemos que para que $2^{11} - 1$ sea divisible entre p es necesario que el d correspondiente a p divida a 11, pero como d no puede ser 1 (esto es evidente), ha de ser $d = 11$, y por el teorema de Fermat $11 \mid p - 1$. En resumen, para que un primo p pueda dividir a $2^{11} - 1$ es necesario que sea de la forma $p = 11m + 1$.

Buscamos ahora los números de la forma $11m + 1$ menores que 45. Podemos ahorrarnos los valores impares de m porque dan números pares, que no pueden ser primos. Resulta que sólo hay un valor posible: $11 \cdot 2 + 1 = 23$. Acabamos de probar que si $2^{11} - 1$ no es primo, entonces es divisible entre 23.

Ahora calculamos módulo 23:

$$\begin{aligned}
\bar{2}^5 &= \overline{32} = \bar{9},\\
\bar{2}^{10} &= \bar{9}^2 = \bar{9} \cdot \bar{3} \cdot \bar{3} = \overline{27} \cdot \bar{3} = \bar{4} \cdot \bar{3} = \overline{12},\\
\bar{2}^{11} &= \overline{24} = \bar{1}.
\end{aligned}$$

Luego resulta que, después de todo, M_{11} no es primo. Concretamente factoriza como $23 \cdot 89$.

Notemos que, en general, hemos demostrado lo siguiente:

> *Si p es un número primo, todo divisor primo de $M_p = 2^p - 1$ es de la forma $q = 2pk + 1$.*

El argumento es que $2^p \equiv 1 \pmod{q}$, por lo que el teorema de Fermat (cuya demostración tenemos pendiente) implica que el menor d que cumple $\bar{2}^d \equiv 1 \pmod{q}$ cumple $d \mid p$ y $d \mid q - 1$, luego es $d = p$, luego $q = pk + 1$, pero k tiene que ser par, luego de hecho $q = 2pk + 1$.

Vamos a aplicar esto al siguiente número de Mersenne, $M_{13} = 2^{13} - 1 = 8\,191$. Así veremos la potencia de la técnica que hemos desarrollado.

En primer lugar $2^{13}-1 < 2^{14} = \left(2^7\right)^2 = 128^2$, luego nos basta buscar posibles divisores primos menores que 128. Acabamos de ver que si $q \mid 2^{13}-1$ es primo, necesariamente $q = 26n+1$.

Los valores de $26n+1$ menores que 128 son $27, \mathbf{53}, \mathbf{79}, 105$. Sólo hay dos primos, lo que significa que si $2^{13}-1$ no es primo entonces es divisible entre 53 o 79. Es fácil ver que no es el caso. Tomamos clases módulo 53:

$$\bar{2}^6 = \overline{64} = \overline{11}, \quad \bar{2}^{12} = \overline{11}^2 = \overline{121} = \overline{15}, \quad \bar{2}^{13} = \overline{30} \neq \bar{1}.$$

Y ahora módulo 79:

$$\begin{array}{lcll} \bar{2}^6 & = & \overline{64} & = -\overline{15}, \\ \bar{2}^{12} & = & -\overline{15}^2 & = -\overline{15}\cdot(-\bar{5})\cdot\bar{3} = \overline{75}\cdot\bar{3} = -\bar{4}\cdot\bar{3} = -\overline{12}, \\ \bar{2}^{13} & = & -\overline{24} & \neq \bar{1}. \end{array}$$

Así pues, M_{13} es primo.

Ejercicio: Probar que $n-1 \mid n^k-1$, con lo que si $k>1$ el número n^k-1 no puede ser primo salvo si $n=2$.

Tenemos pendiente demostrar el teorema de Fermat que hemos usado en el análisis de la primalidad de los números de Mersenne. La prueba es muy simple, pero nos ocupamos de ella en la sección siguiente, donde probaremos un resultado mucho más general.

3.6 Las unidades de $\mathbb{Z}_n$

En la sección anterior hemos usado que si p es primo y $p \nmid a$, entonces existe un mínimo número natural $d>0$ tal que $a^d \equiv 1 \pmod{p}$, así como que $d \mid p-1$. Vamos a tratar de encontrar un resultado más general válido para un módulo m arbitrario, no necesariamente primo.

Veamos primero qué sucede, por ejemplo, si tomamos $m=21$. La tabla muestra, para cada valor de a, el mínimo natural $d>1$ tal que $a^d \equiv 1 \pmod{21}$ cuando existe tal número (no hemos incluido $a=0$, pero es obvio que en tal caso no existe d):

a	1	2	3	4	5	6	7	8	9	10
d	1	6	–	3	6	–	–	2	–	6
a	11	12	13	14	15	16	17	18	19	20
d	6	–	2	–	–	3	6	–	6	2

Por ejemplo, la sucesión de potencias de $\bar{5}$ es:

$$\bar{5}, \quad \bar{4}, \quad -\bar{1}, \quad -\bar{5}, \quad -\bar{4}, \quad \bar{1}, \quad \bar{5}, \quad \bar{4}, \quad \ldots$$

que toma el valor $\bar{1}$ en los exponentes múltiplos de 6, mientras que la sucesión de potencias de 6 es:

$$\bar{6}, \quad -\bar{6}, \quad \bar{6}, \quad -\bar{6}, \quad \ldots$$

que nunca toma el valor $\bar{1}$.

Resumen 3.2: Definición de grupo

Un *grupo* es un conjunto en el que hay definido un producto que cumple las propiedades siguientes:

Asociativa $(ab)c = a(bc)$.

Elemento neutro Existe un elemento 1 tal que $a \cdot 1 = 1 \cdot a = a$.

Elemento simétrico Para cada a existe un a^{-1} tal que $aa^{-1} = a^{-1}a = 1$.

El grupo es *abeliano* si además cumple la propiedad:

Conmutativa $ab = ba$.

¿Puede el lector conjeturar o razonar para qué valores de a existe d? No es cierto que $d \mid 20 = 21 - 1$, pero, vemos que siempre $d \mid 12$. ¿Sabría el lector interpretar ese 12?

Ejercicio: Construir tablas similares, por ejemplo para $m = 22$ y tratar de obtener una conjetura que generalice al teorema de Fermat.

Es fácil ver que, para que pueda darse la congruencia $a^d \equiv 1 \pmod{m}$, con $d > 0$, es necesario que $\bar{a}$ sea una unidad de $\mathbb{Z}_m$. Si $d = 1$ tenemos que $\bar{a} = \bar{1}$ es trivialmente una unidad y, si $d > 1$, entonces $\bar{a} \cdot \bar{a}^{d-1} = \bar{1}$, luego $\bar{a}$ es una unidad y su inversa es $\bar{a}^{-1} = \bar{a}^{d-1}$.

En la tabla anterior vemos que d está definido precisamente para las unidades de $\mathbb{Z}_{21}$, es decir, para las ¡12! clases de restos primos con 21. Vamos a probar que, en general, si $\bar{a}$ es una unidad de $\mathbb{Z}_m$, entonces existe un número $k > 0$ tal que $a^k \equiv 1 \pmod{m}$.

En realidad este resultado (debido a Euler) es un caso particular de otro mucho más general, y en lo sucesivo sacaremos mucho partido a dicha generalidad. Para enunciarlo tenemos que introducir el concepto de *grupo,* que es un conjunto dotado de una operación que cumple las propiedades indicadas en el resumen 3.2.

Es inmediato que el conjunto de las unidades de un dominio constituye un grupo abeliano. En efecto, si a y b son unidades, entonces ab es una unidad porque admite como inverso a $a^{-1}b^{-1}$, luego el producto del dominio se restringe a un producto en el conjunto de sus unidades. Dicho producto es asociativo y conmutativo, porque así lo exige la definición de dominio y, evidentemente, 1 es el elemento neutro y toda unidad tiene un inverso por definición de unidad.

Definición 3.18 Si $m \geq 2$ es un número natural, llamaremos U_m al grupo de las unidades de $\mathbb{Z}_m$, que está formado por las clases de restos $\bar{a}$ tales que $(a, m) = 1$. Así U_m es un grupo abeliano finito.

El número de elementos de un grupo finito G recibe el nombre de *orden* del grupo, y se representa por $|G|$.

Si g es un elemento de un grupo y m un número entero, podemos definir las potencias de g de la forma obvia:

$$g^m = \begin{cases} \overbrace{g \cdots g}^{m \text{ veces}} & \text{si } m > 0 \\ 1 & \text{si } m = 0, \\ \underbrace{g^{-1} \cdots g^{-1}}_{-m \text{ veces}} & \text{si } m < 0. \end{cases}$$

Es fácil ver que $g^{m+n} = g^m g^n$ y $(g^m)^n = g^{mn}$. Además, si dos elementos g y h conmutan, es decir, si $gh = hg$, entonces $(gh)^m = g^m h^n$.

Se dice que g tiene *orden finito* si existe un $m > 0$ tal que $g^m = 1$. En tal caso se define el *orden* de g como el menor número natural $m > 0$ que cumple esto, y se representa por $o(g)$. En caso contrario se dice que g tiene *orden infinito,* y se representa por $o(g) = \infty$.

Teorema 3.19 *Sea g un elemento de un grupo. Entonces:*

1. *Si $o(g) = d$, entonces $g^m = g^n$ si y sólo si $m \equiv n$ (mód d). En particular, $g^m = 1$ si y sólo si $d \mid m$.*
2. *Si $o(g) = \infty$, entonces $g^m = g^n$ si y sólo si $m = n$.*

Demostración: 1) Si $m = dc + r$, con $0 \leq r < d$, se cumple que

$$g^m = (g^d)^c g^r = g^r$$

y, como d es el menor natural no nulo tal que $g^d = 1$, se cumplirá $g^m = 1$ si y sólo si $r = 0$, es decir, si y sólo si $d \mid m$.

Más en general, se cumple $g^m = g^n$ si y sólo si $g^{m-n} = 1$, si y sólo si $d \mid m - n$, si y sólo si $m \equiv n$ (mód d).

2) Si $g^m = g^n$, con $n \leq m$, entonces $g^{m-n} = 1$, y tiene que ser $m - n = 0$ o, de lo contrario, el orden de g sería finito. ∎

Observemos que lo que expresa este teorema es que si un elemento tiene orden finito d, entonces, al calcular las potencias sucesivas de d obtenemos un ciclo en el que aparecen exactamente d elementos del grupo, empezando en $1 = g^0$, y que vuelve a repetirse al llegar a $g^d = 1$. En cambio, si el orden es infinito, no hay potencias repetidas.

Por ejemplo, hemos visto que en el grupo de unidades U_{21} la clase $\bar{5}$ tiene orden 6, pues la sucesión de sus potencias es

n	0	1	2	3	4	5	6	7	8	$\cdots$
$\bar{5}^n$	$\bar{1}$	$\bar{5}$	$\bar{4}$	$-\bar{1}$	$-\bar{5}$	$-\bar{4}$	$\bar{1}$	$\bar{5}$	$\bar{4}$	$\cdots$

Definición 3.20 Si $(a, m) = 1$, definimos el *orden* de a módulo m como el orden de la clase de restos $\bar{a}$ en el grupo U_m, es decir, como el menor número natural $o_m(a) = d > 0$ tal que $a^d \equiv 1$ (mód m).

En estos términos, $o_{21}(5) = 6$. Ahora probamos el resultado fundamental sobre órdenes, que es una versión abstracta, mucho más general, del teorema de Fermat. En realidad es válido también para grupos no abelianos, pero en el caso abeliano el argumento es mucho más simple:

Teorema 3.21 *Si G es un grupo abeliano finito y g es un elemento de G, entonces g tiene orden finito y $o(g) \mid |G|$.*

DEMOSTRACIÓN: Sea $n = |G|$ y sean $g_1, \ldots, g_n$ todos los elementos de G. Entonces $gg_1, \ldots, gg_n$ son también todos los elementos de g, aunque tal vez en otro orden. En efecto, observemos que si $gg_i = gg_j$, multiplicando por g^{-1} concluimos que $g_i = g_j$, luego los elementos gg_i son n elementos de G distintos dos a dos, pero como G tiene exactamente n elementos, son todos los elementos de G. Por consiguiente:

$$(gg_1)\cdots(gg_n) = g_1 \cdots g_n,$$

ya que se trata de los mismos elementos multiplicados en un orden distinto (y estamos suponiendo que G es abeliano). Reagrupando los términos de la izquierda tenemos que

$$g^n g_1 \cdots g_n = g_1 \cdots g_n.$$

Multiplicando por $(g_1 \cdots g_n)^{-1}$ concluimos que $g^n = 1$. Esto prueba que g tiene orden finito, así como que $o(g) \mid n$. ∎

Para particularizar este resultado al caso de U_n conviene introducir la función de Euler:

Definición 3.22 La *función de Euler* es la función que a cada número natural $m \geq 2$ le asigna el orden $\phi(m)$ del grupo de unidades U_m o, alternativamente, el número de números $1 \leq a < m$ tales que $(a, m) = 1$. Convenimos además en que $\phi(1) = 1$.

Teorema 3.23 (Euler) *Si $m \geq 2$ y $(a, m) = 1$, entonces $o_m(a) \mid \phi(m)$.*

Este resultado incluye como caso particular el teorema de Fermat que hemos usado en la sección anterior, pues si p es primo, es claro que $\phi(p) = p - 1$, luego en este caso lo que tenemos es que si $p \nmid a$, entonces $o_p(a) \mid p - 1$.

Teorema 3.24 (Fermat) *Si p es primo, y n es un entero tal que $p \nmid n$, entonces $o_p(n) \mid p - 1$. En particular $n^{p-1} \equiv 1$ (mód p).*

Ahora entendemos por qué los órdenes de las clases de U_{21} son divisores de 12. Ese 12 es precisamente $\phi(21) = 12$.

Calcular $\phi(n)$ a partir de la definición de la función de Euler es bastante laborioso, pero hay una forma mucho más sencilla de hacerlo. El lector podría tratar de conjeturarla a partir de la tabla siguiente, que muestra los primeros valores de $\phi(m)$:

1 1	2 1	3 2	4 2	5 4	6 2	7 6	8 4	9 6	10 4
11 10	12 4	13 12	14 6	15 8	16 8	17 16	18 6	19 18	20 8
21 12	22 10	23 22	24 8	25 20	26 12	27 18	28 12	29 28	30 8
31 30	32 16	33 20	34 16	35 24	36 12	37 36	38 18	39 24	40 16
41 40	42 12	43 42	44 20	45 24	46 22	47 46	48 16	49 42	50 20
51 32	52 24	53 52	54 18	55 40	56 24	57 36	58 28	59 58	60 16
61 60	62 30	63 36	64 32	65 48	66 20	67 66	68 32	69 44	70 24
71 70	72 24	73 72	74 36	75 40	76 36	77 60	78 24	79 78	80 32
81 54	82 40	83 82	84 24	85 64	86 42	87 56	88 40	89 88	90 24
91 72	92 44	93 60	94 46	95 72	96 32	97 96	98 42	99 60	100 40

Si el ojo del lector ya está suficientemente entrenado, habrá advertido que ϕ es multiplicativa. Esto es una consecuencia del teorema chino del resto. Consideremos, por ejemplo, el caso de $\phi(15) = \phi(3)\phi(5)$.

La tabla siguiente muestra las clases de $\mathbb{Z}_{15}$ (menos $\bar{0}$) y sus representaciones como pares de $\mathbb{Z}_3 \times \mathbb{Z}_5$:

$\mathbf{\bar{1}}$	$\mathbf{\bar{2}}$	$\bar{3}$	$\mathbf{\bar{4}}$	$\bar{5}$	$\bar{6}$	$\mathbf{\bar{7}}$
$(\bar{1},\bar{1})$	$(\bar{2},\bar{2})$	$(\bar{0},\bar{3})$	$(\bar{1},\bar{4})$	$(\bar{2},\bar{0})$	$(\bar{0},\bar{1})$	$(\bar{1},\bar{2})$
$\mathbf{\bar{8}}$	$\bar{9}$	$\mathbf{\bar{10}}$	$\mathbf{\bar{11}}$	$\bar{12}$	$\mathbf{\bar{13}}$	$\mathbf{\bar{14}}$
$(\bar{2},\bar{3})$	$(\bar{0},\bar{4})$	$(\bar{1},\bar{0})$	$(\bar{2},\bar{1})$	$(\bar{0},\bar{2})$	$(\bar{1},\bar{3})$	$(\bar{2},\bar{4})$

Vemos que las unidades se corresponden con pares de unidades, y esto no es casual. Si $\bar{a}$ es una unidad y se corresponde con el par $(\bar{c},\bar{d})$, entonces $\bar{a}^{-1}$ se corresponderá con un par $(\bar{u},\bar{v})$ de modo que

$$(\bar{c},\bar{d})(\bar{u},\bar{v}) = (\bar{c}\,\bar{u},\bar{d}\,\bar{v}) = (\bar{1},\bar{1}),$$

luego $\bar{c}$ y $\bar{d}$ serán unidades de sus anillos respectivos. Recíprocamente, si $\bar{a}$ se corresponde con un par de unidades $(\bar{c},\bar{d})$, entonces el par $(\bar{c}^{-1},\bar{d}^{-1})$ se corresponde con un $\bar{b}$ que cumplirá $\bar{a}\,\bar{b} = \bar{1}$, luego $\bar{a}$ será una unidad.

Este argumento general muestra que, si $(m,n) = 1$, el número de unidades de $\mathbb{Z}_{mn}$ es el producto del número de unidades de $\mathbb{Z}_m$ por el número de unidades de $\mathbb{Z}_n$. Esto es justo lo que expresa la igualdad $\phi(mn) = \phi(m)\phi(n)$.

Por consiguiente, para calcular fácilmente la función de Euler basta encontrar una fórmula para $\phi(p^n)$, cuando n es primo. La solución es:

$$\phi(p^n) = (p-1)p^{n-1}.$$

En efecto, números menores que p^n hay $p^n - 1$, de los cuales no serán primos con p^n los múltiplos de p, es decir, los de la forma $pk < p^n$, de modo que $k < p^{n-1}$, y en total hay $p^{n-1} - 1$, luego serán primos con p^n los restantes:

$$p^n - 1 - (p^{n-1} - 1) = p^n - p^{n-1} = (p-1)p^{n-1}.$$

Por ejemplo,

$$\phi(100) = \phi(2^2)\phi(5^2) = (2-1)2^{2-1}(5-1)5^{2-1} = 40.$$

Ejercicio: Gauss demostró que el polígono regular de n lados es constructible con regla y compás si y sólo si $\phi(n)$ es potencia de 2. Caracterizar los números que cumplen esta propiedad en términos de su descomposición en factores primos. Encontrar (sin calculadora) un n impar con esta propiedad que sea mayor que 50 000.

Nota En los capítulos precedentes hemos demostrado varios teoremas no triviales de "álgebra abstracta", como la construcción de los cuerpos de cocientes, el hecho de que todo dominio euclídeo es un dominio de factorización única, así como todos los teoremas generales que hemos probado sobre dominios de factorización única. Todos ellos son "abstractos" en el sentido de que se prueban para dominios "genéricos", de forma que podemos particularizarlos a casos muy diversos. Por ejemplo, el mismo teorema general nos justifica que $\mathbb{Z}$ y $\mathbb{Q}[x]$ tienen factorización única.

Sin embargo, todos ellos, sin ser triviales, pueden considerarse "abstracciones triviales" en el sentido de que los argumentos necesarios para probarlos en toda su generalidad son prácticamente los mismos que habría que emplear para probarlos en un caso particular. Son una mera constatación de que un argumento es válido en un contexto más general que el primero en que uno puede considerarlo. No hacen más que evitar que tengamos que repetir varias veces un mismo argumento en contextos distintos.

Por el contrario, el teorema 3.21 y, en general, los elementos de teoría de grupos que hemos presentado en esta sección, son diferentes en este sentido. Uno puede entender perfectamente la prueba de que $\mathbb{Z}$ o $\mathbb{Q}[x]$ tienen factorización única sin echar de menos para nada el concepto abstracto de dominio de factorización única, pero, en cambio, los argumentos expuestos en esta sección "se entienden más claramente" cuando se plantean en términos abstractos que cuando nos los encontramos en casos particulares mezclados con hechos accidentales que resultan ser irrelevantes (como el hecho de estar considerando clases de congruencias, etc.).

Por ejemplo, en el apartado siguiente vamos a demostrar un resultado que Euler conjeturó, pero no supo probar (la primera demostración es de Gauss) y veremos que las ideas involucradas resultan muy simples precisamente gracias a que pueden ser expresadas en los términos abstractos de la teoría de grupos y cuerpos. ∎

Raíces primitivas Consideremos los órdenes de las distintas clases de restos respecto de un módulo primo, por ejemplo, $p = 11$:

a	1	2	3	4	5	6	7	8	9	10
$o_{11}(a)$	1	10	5	5	5	10	10	10	5	2

El teorema de Fermat predice que todos los órdenes tienen que ser divisores de 10, pero vemos que en este caso sucede algo más fuerte: hay clases cuyo orden es exactamente 10.

Definición 3.25 Una *raíz primitiva* en U_m (o una raíz primitiva módulo m) es un entero a tal que $(a, m) = 1$ y $o_m(a) = \phi(m)$.

Acabamos de ver que existen raíces primitivas módulo 11, mientras que la tabla que hemos dado al principio de esta sección muestra que no hay raíces primitivas módulo 21.

Las raíces primitivas simplifican mucho los cálculos con unidades, pues, sustituyendo cada clase por su exponente, podemos calcular productos calculando, en realidad, sumas. Más precisamente, si a es una raíz primitiva módulo m y $\phi(m) = d$, podemos definir un índice:

$$\text{índ}_a : U_m \longrightarrow \mathbb{Z}_d$$

de modo que $\text{índ}_a(\bar{n}) = \bar{\imath}$ si y sólo si $\bar{a}^i = \bar{n}$.

Notemos que la definición es correcta, pues si $\bar{a}^i = \bar{a}^j$, entonces $\bar{a}^{i-j} = \bar{1}$, luego $d \mid i - j$, luego $\bar{\imath} = \bar{\jmath}$.

Por ejemplo, si $m = 11$ y $a = 2$, tenemos que las potencias de 2 son:

n	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{7}$	$\bar{8}$	$\bar{9}$
$\bar{2}^n$	$\bar{1}$	$\bar{2}$	$\bar{4}$	$\bar{8}$	$\bar{5}$	$\overline{10}$	$\bar{9}$	$\bar{7}$	$\bar{3}$	$\bar{6}$

por lo que los índices de base 2 son:

n	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{7}$	$\bar{8}$	$\bar{9}$	$\overline{10}$
$\text{índ}_2(n)$	$\bar{0}$	$\bar{1}$	$\bar{8}$	$\bar{2}$	$\bar{4}$	$\bar{9}$	$\bar{7}$	$\bar{3}$	$\bar{6}$	$\bar{5}$

Es inmediato que

$$\text{índ}_a(mn) = \text{índ}_a(m) + \text{índ}_a(n),$$

por lo que si, por ejemplo, queremos calcular $\bar{4}^2 \cdot \bar{7}^3$, es más fácil hacerlo a través de los índices:

$$\text{índ}_2(\bar{4}^2 \cdot \bar{7}^3) = 2 \cdot \text{índ}_2(\bar{4}) + 3 \cdot \text{índ}_2(\bar{7}) = 2 \cdot \bar{2} + 3 \cdot \bar{7} = \bar{4} + \overline{21} = \bar{4} + \bar{1} = \bar{5},$$

(notemos que las congruencias de índices son módulo 10), luego $\bar{4}^2 \cdot \bar{7}^3 = \overline{10}$.

Ejercicio: Si g es un elemento de orden n en un grupo G. Encontrar una expresión para el orden de g^m en función de m y n. Comprobarla / conjeturarla con los elementos de U_{11} tomndo $g = \bar{2}$.

Vamos a probar que existen raíces primitivas módulo cualquier primo p. La prueba no es complicada, pero el hecho no es obvio en absoluto. Se basa en unos resultados elementales de teoría de grupos.

Definición 3.26 Un grupo G es *cíclico* si existe un elemento g en G tal que todos los demás elementos de G son potencias de g. En tal caso se dice que g es un *generador* de G.

Observemos en primer lugar que el hecho de que un grupo U_m contenga una raíz primitiva equivale a que U_m es un grupo cíclico, y sus generadores son precisamente las raíces primitivas.

Notemos que un grupo finito G es cíclico si y sólo si tiene un elemento g de orden $|G|$, pues esto implica que g tiene $|G|$ potencias distintas, luego dichas potencias son todos los elementos de G y, recíprocamente, si G es cíclico, el orden de un generador tiene que ser $|G|$.

También es obvio que todo grupo cíclico es abeliano, pues todos sus elementos son de la forma g^n y

$$g^n g^m = g^{n+m} = g^{m+n} = g^m g^n.$$

Teorema 3.27 *Sea G un grupo y sean g, h dos elementos tales que $gh = hg$ y que tengan órdenes finitos primos entre sí. Entonces gh tiene orden finito y $o(gh) = o(g)o(h)$.*

DEMOSTRACIÓN: Sean $m = o(g)$, $n = o(h)$. Como $gh = hg$, es fácil ver que

$$(gh)^{mn} = g^{mn}h^{mn} = (g^m)^n(h^n)^m = 1^n 1^n = 1.$$

Esto prueba que gh tiene orden finito y además $o(gh) \mid mn$. Claramente, podemos expresar $o(gh) = m_1 n_1$, donde $m_1 \mid m$, $n_1 \mid n$. Tenemos entonces que

$$a^{m_1 n_1} b^{m_1 n_1} = (ab)^{m_1 n_1} = 1.$$

Elevando a m/m_1 queda

$$a^{m n_1} b^{m n_1} = 1,$$

pero $a^m = 1$, luego $b^{m n_2} = 1$, luego $n \mid m n_1$, pero $(n, m) = 1$, luego $n \mid n_1$ y así $n = n_1$. Igualmente se razona que $m = m_1$. ∎

Teorema 3.28 *En todo grupo abeliano finito existe un elemento cuyo orden es múltiplo de los órdenes de todos los demás elementos del grupo.*

DEMOSTRACIÓN: Sea g un elemento del grupo cuyo orden sea máximo. Sea h cualquier otro elemento y vamos a ver que $o(h) \mid o(g)$. Sea $o(g) = m$, $o(h) = n$. Pongamos que

$$m = p_1^{e_1} \cdots p_r^{e_r}, \qquad n = p_1^{e_1'} \cdots p_r^{e_r'},$$

sea m' el producto de las potencias $p_i^{e_i}$ tales que $e_i \geq e_i'$ y sea n' el producto de las potencias $p_i^{e_i'}$ tales que $e_i' > e_i$. De este modo $m' \mid m$, $n' \mid n$, $(m', n') = 1$ y $m'n'$ es el mínimo común múltiplo de m y n.

Sean $g' = g^{m/m'}$ y $h' = h^{n/n'}$. Se cumple que $o(g') = m'$ y $o(h') = n'$. En efecto, por una parte $g'^{m'} = g^m = 1$ y, si $g'^k = 1$, entonces $g^{km/m'} = 1$, luego $m \mid km/m'$, luego $m'm \mid km$, luego $m' \mid k$.

Esto prueba que $o(g') = m'$, e igualmente se razona con h'. Por el teorema anterior $o(g'h') = m'n'$, que es el mínimo común múltiplo de m y n. Teniendo en cuenta que m es el máximo orden de un elemento del grupo, resulta que $m \leq m'n' \leq m$, luego $m = m'n'$, luego $n \mid m$. ∎

Ahora ya podemos probar un resultado mucho más general que la existencia de raíces primitivas en los grupos U_p:

Teorema 3.29 *Si k es un cuerpo finito, su grupo de unidades es cíclico.*

DEMOSTRACIÓN: Sea U el grupo de unidades de k (que consta de hecho de todos los elementos de k menos 0). Por el teorema anterior existe una unidad u en U cuyo orden m es múltiplo de los órdenes de todas las unidades de k o, equivalentemente, tal que $a^m = 1$ para toda unidad a de U. Esto se interpreta como que todos los elementos de U son raíces del polinomio $x^m - 1$ de $k[x]$, pero un polinomio de grado m tiene a lo sumo m raíces, luego $|U| \leq m$.

Por otra parte, las u tiene m potencias distintas, luego U tiene al menos m elementos, así que $|U| = m$ y los elementos de U son precisamente las potencias de u. Esto prueba que u es un generador de U. ∎

En particular, si p es primo, el grupo U_p es cíclico, es decir:

Teorema 3.30 *Si p es primo, existen raíces primitivas módulo p.*

Ejercicio: Encontrar raíces primitivas módulo los primeros primos.

Teorema 3.31 *Si p es un primo impar y g es una raíz primitiva módulo p, entonces o bien g o bien $g + p$ es una raíz primitiva módulo p^2.*

DEMOSTRACIÓN: Sabemos que $\phi(p^2) = (p-1)p$. Sea $m = o_{p^2}(g)$. Entonces $p^2 \mid g^m - 1$, luego en particular $g^m \equiv 1 \pmod{p}$, luego $p-1 = o_p(g) \mid m \mid p(p-1)$. Esto sólo deja dos posibilidades: o bien $m = p-1$, o bien $m = (p-1)p$, y en el segundo caso g es una raíz primitiva módulo p^2.

Lo mismo se aplica a $g + p$, que obviamente es una raíz primitiva módulo p (ya que determina la misma clase). Basta probar que no puede ocurrir que $o_{p^2}(g) = o_{p^2}(g + p) = p - 1$.

En tal caso $g^{p-1} \equiv 1 \pmod{p^2}$, luego $g^p \equiv g \pmod{p^2}$, luego

$$g \equiv g^p \equiv g^p + p^p \equiv (g+p)^p \equiv (g+p)^{p-1}(g+p) \equiv g + p \pmod{p^2},$$

de donde $p \equiv 0 \pmod{p^2}$, lo cual es absurdo. ∎

Ejercicio: Encontrar raíces primitivas módulo p^2 para los primeros primos. ¿Existen raíces primitivas módulo 4?

Ejercicio: Comprobar que 14 es una raíz primitiva módulo 29, pero no módulo 29^2.

Ejercicio: ¿Cuántos elementos de orden n tiene un grupo cíclico de orden n? En particular, ¿cuántas raíces primitivas hay en U_{p^n}?

Teorema 3.32 *Si p es un primo impar y g es una raíz primitiva módulo p^2, entonces es una raíz primitiva módulo p^n, para todo n.*

DEMOSTRACIÓN: Recordemos que $\phi(p^n) = (p-1)p^{n-1}$. Tenemos que

$$g^{p-1} \equiv 1 \text{ (mód } p), \qquad g^{p-1} \not\equiv 1 \text{ (mód } p^2),$$

luego $v_p(g^{p-1} - 1) = 1$. Supuesto que g sea una raíz primitiva módulo p^n, llamemos $m = o_{p^{n+1}}(g)$, de modo que $g^m \equiv 1$ (mód p^{n+1}), luego en particular $g^m \equiv 1$ (mód p^n), luego $(p-1)p^{n-1} \mid m \mid (p-1)p^n$. Ahora bien, el teorema 3.6 nos da que

$$v_p(g^{(p-1)p^{n-1}} - 1) = v_p(g^{(p-1)p^{n-1}} - 1^{(p-1)p^{n-1}}) = v_p(g^{p-1} - 1^{p-1}) + n - 1 = n,$$

luego $p^{n+1} \nmid g^{(p-1)p^{n-1}} - 1$, es decir, que $g^{(p-1)p^{n-1}} \not\equiv 1$ (mód p^{n+1}), luego $m = (p-1)p^n$, y esto prueba que g es una raíz primitiva módulo p^{n+1}. ∎

Ejercicio: Estudiar si existen raíces primitivas módulo n para $2 \le n \le 16$.

Teorema 3.33 *Si $n \ge 3$, entonces $|U_{2^n}| = 2^{n-1}$ y $o_{2^n}(\bar{5}) = 2^{n-2}$. Toda clase de U_{2^n} se expresa de forma única como $\pm\bar{5}^j$, con $j = 0, \ldots, 2^{n-2} - 1$.*

DEMOSTRACIÓN: Veamos que, para $n \ge 3$,

$$5^{2^{n-3}} \equiv 1 + 2^{n-1} \text{ (mód } 2^n).$$

En efecto, se comprueba que es cierto para $n = 3$ y, si vale para un n arbitrario, entonces $5^{2^{n-3}} = 1 + 2^{n-1} + k2^n$, luego

$$5^{2^{n+1-3}} = (5^{2^{n-3}})^2 = 1 + 2^n + k2^{n+1} + 2^{2n-2} + k^2 2^{2n} + k2^{2n},$$

pero $n + 1 \le 2n - 2 \le 2n$, luego $5^{2^{n+1-3}} \equiv 1 + 2^n$ (mód 2^{n+1}).

Por lo tanto, $\bar{5}^{2^{n-3}} \ne \bar{1}$ en U_{2^n}. En cambio,

$$\bar{5}^{2^{n-2}} = (\bar{5}^{2^{n-1}})^2 = (1 + \bar{2}^{n-1})^2 = \bar{1}.$$

Así pues, $o_{2^n}(5) \mid 2^{n-2}$, pero $o_{2^n}(5) \nmid 2^{n-3}$, lo que implica que $o_{2^n}(5) = 2^{n-2}$.

Veamos ahora que si $(-\bar{1})^i\bar{5}^j = (-\bar{1})^{i'}\bar{5}^{j'}$, donde $i, i' = 0, 1$, $0 \le j \le j' < 2^{n-2}$, necesariamente $i = i'$, $j = j'$. En efecto, tenemos que

$$(-\bar{1})^{i-i'} = \bar{5}^{j'-j}.$$

Si fuera $i \ne i'$, tendríamos $\bar{5}^{j'-j} + \bar{1} = \bar{0}$, es decir, que $2^n \mid 5^{j-j'} + 1$. En particular $4 \mid 5^{j-j'} + 1$, pero $5^{j-j'} + 1 \equiv 1 + 1 \equiv 2$ (mód 4), y tenemos una contradicción, luego tiene que ser $i = i'$, luego $\bar{5}^{j'-j} = \bar{0}$, pero el orden de $\bar{5}$ es mayor que $j' - j$, luego $j = j'$.

Así pues, las clases de la forma $\pm\bar{5}^j$ son 2^{n-1} clases distintas, luego son todas las clases de U_{2^n}. ∎

Ahora es fácil ver que todos los elementos de U_{2^n} tienen orden divisor de 2^{n-2}, por lo que no hay raíces primitivas módulo 2^n (en principio si $n \geq 3$, pero es fácil ver que tampoco las hay si $n = 2$, y el caso $n = 1$ es trivial).

Veamos un ejemplo en el que el uso de índices simplifica considerablemente un cálculo (o al menos lo simplificaba en los tiempos en los que no había ordenadores, pero sí tablas de índices precalculadas):

Ejemplo Probar que la ecuación $x^2 + 1\,848y^2 = 18\,518\,809$ tiene como única solución (en los números naturales)

$$197^2 + 1\,848 \cdot 100^2 = 18\,518\,809.$$

Conceptualmente, la comprobación no tiene ninguna dificultad. Es fácil ver (sin necesidad siquiera de hacer el cálculo) que $1\,848 \cdot 101^2 > 18\,518\,809$, por lo que otra posible solución de la ecuación debería cumplir $y < 100$, luego sólo se trata de comprobar que los números $18\,518\,809 - 1\,848y^2$ no son cuadrados perfectos para ningún $y < 100$. Ahora bien, calcular 100 raíces cuadradas de números de esta magnitud no es muy agradable.

Para ello observamos que $1\,848 = 2^2 \cdot 3 \cdot 7 \cdot 11$, así que vamos a considerar primos distintos de éstos. Empezamos con $p = 5$ y observamos que la ecuación se reduce a

$$x^2 + 3y^2 \equiv 4 \text{ (mód 5)}.$$

Los valores posibles para x^2 son $x^2 \equiv 0, 1, 4$ (mód 5), luego $3y^2 \equiv 0, 3, 5$ (mód 5). Como $2 \cdot 3 \equiv 1$ (mód 5), multiplicando por 2 llegamos a que $y^2 \equiv 0, 1, 3$ (mód 5), pero 3 no es un cuadrado módulo 5 y, descartándolo, $y \equiv 0, 1, 4$ (mód 5) o, más convenientemente, $y \not\equiv 2, 3$ (mód 5).

Con esto nos quedan sólo 60 valores posibles para y. Podemos determinarlos mediante una criba:

0	1	×	×	4	5	6	×	×	9
10	11	×	×	14	15	16	×	×	19
20	21	×	×	24	25	26	×	×	29
30	31	×	×	34	35	36	×	×	39
40	41	×	×	44	45	46	×	×	49
50	51	×	×	54	55	56	×	×	59
60	61	×	×	64	65	66	×	×	69
70	71	×	×	74	75	76	×	×	79
80	81	×	×	84	85	86	×	×	89
90	91	×	×	94	95	96	×	×	99

Ahora repetimos el proceso con $p = 13$, con lo que la ecuación se reduce a

$$x^2 + 2y^2 \equiv 10 \text{ (mód 13)}.$$

En este punto se agradece disponer de una tabla de índices módulo 13 (por ejemplo, respecto de la raíz primitiva $a = 2$):

n	1	2	3	4	5	6	7	8	9	10	11	12
$\text{ind}_2(n)$	0	1	4	2	9	5	11	3	8	10	7	6

Por ejemplo, los cuadrados los los restos de índice par (más el 0), luego vemos que

$$x^2 \equiv 0, 1, 3, 4, 9, 10, 12 \text{ (mód 13)}.$$

Por lo tanto,

$$2y^2 \equiv 0, 1, 6, 7, 9, 10, 11 \text{ (mód 13)}.$$

La tabla nos da también el inverso de 2 módulo 13. Basta ver que $\text{ind}(2) = 1$, luego su inverso es el resto de índice $-1 \equiv 11$ (mód 12), es decir, 7. Multiplicando por 7 obtenemos

$$y^2 \equiv 0, 3, 5, 7, 10, 11, 12 \text{ (mód 13)}.$$

Las multiplicaciones también pueden hacerse mediante la tabla. Para ello observamos que los índices de $1, 6, 7, 9, 10, 11$ son $0, 5, 11, 8, 10, 7$, respectivamente, y sólo tenemos que sumarles $\text{ind}(7) = -1$, con lo que obtenemos $11, 4, 10, 7, 9, 6$, que son los índices de los valores posibles para y^2 (aparte del 0). En realidad no necesitamos pasar a los restos, sino que podemos seguir trabajando con los índices y descartar los impares, porque no corresponden a cuadrados. Así,

$$\text{ind}(y^2) \equiv 4, 6, 10 \text{ (mód 12)}$$

Cada valor para y^2 se corresponde con dos valores para y, el que resulta de dividir el índice entre 2 y en que resulta de sumarle 6, pues $2(a + 6) \equiv 2a$ (mód 12), luego los restos de índice a y $a + 6$ tienen el mismo cuadrado. El resultado es que

$$\text{ind}(y) \equiv 2, 3, 5, 8, 9, 11 \text{ (mód 12)},$$

lo que se corresponde con que

$$y \equiv 0, 4, 5, 6, 7, 8, 9 \text{ (mód 13)}$$

o, más cómodamente, $y \not\equiv 1, 2, 3, 10, 11, 12$ (mód 13). Al realizar la criba correspondiente sobreviven 33 valores para y:

0	×	×	×	4	5	6	×	×	9
×	×	×	×	×	×	×	×	×	19
20	21	×	×	×	×	26	×	×	×
30	31	×	×	34	35	×	×	×	39
×	×	×	×	44	45	46	×	×	×
×	×	×	×	×	×	56	×	×	59
60	61	×	×	×	65	×	×	×	69
70	71	×	×	74	×	×	×	×	×
×	×	×	×	84	85	86	×	×	×
×	91	×	×	×	95	96	×	×	99

El lector puede probar a hacer los cálculos con $p = 17$ con la tabla (para $a = 3$):

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\text{ind}_3(n)$	0	14	1	12	5	15	11	10	2	3	7	13	4	9	6	8

La ecuación es $x^2 + 12y^2 \equiv 2 \pmod{17}$. Encontrará que 17 valores de y sobreviven a la criba. Con $p = 19$ sobreviven 9, mientras que con $p = 23$ las posibilidades se reducen a $y = 39, 84, 86$ (aparte de $y = 100$).

A partir de aquí no compensa hacer los cálculos en general, y vale más tratar con cada posibilidad individualmente. Por ejemplo, con $p = 29$, si tomamos $y = 86$ la ecuación se reduce a $x^2 \equiv 26 \pmod{29}$, pero 26 tiene índice impar, luego esto es imposible. Igualmente se descarta $y = 84$. Por último, con $p = 31$ se descarta el 39.

Este resultado es interesante porque Euler demostró (teorema 14.16) que esto implica que 18 518 809 es primo. Una comprobación directa requeriría probar a dividirlo entre los primeros 590 primos, hasta 4 297. ∎

Ejercicio: Probar que la ecuación $x^2 + 357y^2 = 142\,969$ tiene como única solución (en los números naturales) $(x, y) = (13, 20)$.

Ejemplo Encontrar los números naturales N tales que el número que resulta de trasladar su primera cifra por la izquierda a la derecha (p.ej. $3\,458 \to 4\,583$) es 1.5 veces[10] mayor que N.

Pongamos que un número N en las condiciones descritas tiene n cifras y sea a la primera de ellas (por la izquierda). Entonces $N = a \cdot 10^{n-1} + x$, con $x < 10^{n-1}$ y el número que resulta de pasar a la derecha la primera cifra es $M = 10x + a$. Queremos que se cumpla la relación

$$10x + a = \frac{3}{2}(a \cdot 10^{n-1} + x),$$

que equivale a

$$x = \frac{3 \cdot 10^{n-1} - 2}{17}\, a.$$

Para que se cumpla la condición $x < 10^{n-1}$ tiene que ser

$$\left(3 - \frac{17}{a}\right) \cdot 10^{n-1} < 2$$

y esto, si $n \geq 3$, sólo se cumple si $a \leq 5$. El número N será

$$N = a \cdot 10^{n-1} + \frac{3 \cdot 10^{n-1} - 2}{17}\, a = 2a\left(\frac{10^n - 1}{17}\right)$$

Para que N sea entero n tiene que ser múltiplo de $o_{17}(10) = 16$, luego las soluciones son los números de la forma

$$N_k = 2a\frac{10^{16k} - 1}{17}, \qquad k \geq 1.$$

[10] El lector puede buscar los números que al hacerles dicha transformación duplican al número inicial, pero entonces la segunda cifra por la izquierda resulta ser 0, con lo que el proceso no es reversible.

Observemos además que

$$N_{k+1} - N_k = 2a\frac{10^{16k}}{17} \cdot 10^{16} = N_1 \cdot 10^{16},$$

para $a = 1, 2, 3, 4, 5$. Así pues, N_1 puede ser cualquiera de los cinco números de 16 cifras:

$$1\,176\,470\,588\,235\,294, \qquad 2\,352\,941\,176\,470\,588, \qquad 3\,529\,411\,764\,705\,882,$$

$$4\,705\,882\,352\,941\,176, \qquad 5\,882\,352\,941\,176\,470,$$

y los demás N_k son los que resultan de repetir k veces estos bloques de 16 cifras. En particular, el menor ejemplo que cumple las condiciones es

$$\frac{3}{2} \cdot 1\,176\,470\,588\,235\,294 = 1\,764\,705\,882\,352\,941.$$

■

Multiplicaciones mágicas Este problema fue propuesto en 1907 por Henry Ernest Dudeney:

Éste es un ejemplo de *multiplicación mágica:*

$$\mathbf{41\,096} \times 83 = 3\,\mathbf{410\,96}8.$$

Vemos que el producto se obtiene anteponiendo al primer factor la segunda cifra del segundo y posponiendo la primera.

Encontrar todas las multiplicaciones mágicas, es decir, todos los pares de números naturales (x, y) tales que y tiene dos cifras y el producto xy es el número que resulta de anteponer a x a segunda cifra de y y de posponer la primera.

Pongamos que x tiene n cifras, de modo que $10^{n-1} \le x < 10^n$, mientras que $y = a \cdot 10 + b$, con $a \neq 0$. Notemos que no puede ser $b = 0$, porque entonces el producto xy terminaría en 0, pero tiene que terminar en $a > 0$. La condición es

$$x(a \cdot 10 + b) = b \cdot 10^{n+1} + 10x + a,$$

que equivale a que

$$10^{n-1} \le x = \frac{b \cdot 10^{n+1} + a}{10(a-1) + b} < 10^n.$$

Por lo tanto, para que x sea entero se tiene que cumplir que

$$10(a-1) + b \mid b \cdot 10^{n+1} + a.$$

Es fácil pedirle a un ordenador que compruebe para qué valores de a y b se cumple esta condición para algún n (sólo hay que considerar valores de n hasta $m = 10(a-1) + b$), y sucede que sólo hay tres valores posibles para a y b, es

decir, para y. La tabla siguiente contiene además el menor valor de n posible en cada caso:

x	y	m	n
	71	61	52
41 096	83	73	5
8	86	76	1

El valor de x que falta es un número de 52 cifras, a saber:

$$1\,639\,344\,262\,295\,081\,967\,213\,114\,754\,098\,360\,655\,737\,704\,918\,032\,787.$$

En general, para $y = 71$, los valores posibles de n son los que cumplen

$$10^{n+1} \equiv -7 \text{ (mód } 61),$$

pero $o_{61}(10) = 60$, por lo que esto sucede para $n = 52 + 60k$. Por lo tanto, los valores posibles para x son

$$x_k = \frac{10^{53+60k} + 7}{61},$$

de forma que

$$x_{k+1} - x_k = \frac{10^{53+60k}(10^{60} - 1)}{61} = \frac{10(10^{60} - 1)}{61} \cdot 10^{52+60k},$$

donde el primer factor tiene 60 cifras, y es:

$$163\,934\,426\,229\,508\,196\,721\,311\,475\,409\,836\,065\,573\,770\,491\,803\,278\,688\,524\,590.$$

Esto significa que el factor x_k se obtiene de x_0 añadiendo a su izquierda k bloques de estas 60 cifras, y así tenemos determinadas todas las multiplicaciones mágicas cuyo segundo factor es 71.

Un análisis análogo muestra que, para $y = 83$,

$$x_k = \frac{3 \cdot 10^{6+8k} + 8}{73},$$

con lo que

$$x_{k+1} - x_k = \frac{30(10^8 - 1)}{73} \cdot 10^{5+8k} = 41\,095\,890 \cdot 10^{5+8k},$$

con lo que x_k se obtiene de $x_0 = 41\,096$ añadiendo a su izquierda k bloques iguales a 41 095 890.

El caso $y = 86$ requiere especial atención porque $m = 76$ no es primo. Ahora bien, la condición para n es

$$6 \cdot 10^{n+1} \equiv -8 \text{ (mód } 76),$$

que, dividiendo entre 4, equivale a

$$15 \cdot 10^n \equiv -2 \text{ (mód 19)},$$

o también a $10^n \equiv 10$ (mód 19), y 19 ya es primo. Esta condición se cumple cuando $n = 1 + 18k$, con lo que llegamos a que

$$x_k = \frac{6 \cdot 10^{2+18k} + 8}{76} = \frac{15 \cdot 10^{1+18k} + 2}{19},$$

con lo que

$$x_{k+1} - x_k = \frac{15(10^{18} - 1)}{19} \cdot 10^{1+18k} = 789\,473\,684\,210\,526\,315 \cdot 10^{1+18k},$$

y así x_k se obtiene de $x_0 = 8$ anteponiendo k bloques 789 473 684 210 526 315. Así quedan determinadas todas las multiplicaciones mágicas. La tabla siguiente resume lo que hemos obtenido:

periodo	x_0	y
789 473 684 210 526 315	8	86
41 095 890	41 096	83
163 934 426 229 508 196	1 639 344 262 295 081	71
721 311 475 409 836 065 573	967 213 114 754 098 360	
770 491 803 278 688 524 590	655 737 704 918 032 787	

■

Un teorema de irreducibilidad Veamos ahora una aplicación más sofisticada de las congruencias en la que usaremos buena parte de los resultados que hemos probado hasta ahora:

Teorema 3.34 *Si p es un primo impar, el polinomio*

$$c_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$$

es irreducible en $\mathbb{Q}[x]$.

Demostración: Supongamos que $c_p(x) = f(x)g(x)$, para ciertos polinomios en $\mathbb{Q}[x]$ de grado estrictamente menor que el de $c_p(x)$. Sustituyendo $f(x)$ y $g(x)$ por los polinomios que resultan de multiplicarlos por el producto de los denominadores de sus coeficientes, obtenemos polinomios del mismo grado con coeficientes enteros, que satisfacen una ecuación de la forma

$$ac_p(x) = f(x)g(x),$$

donde a es un entero no nulo, que podemos suponer positivo. Supongamos en primer lugar que $a > 1$ y tomemos un primo $r \mid a$. Así r divide a todos los coeficientes de $f(x)g(x)$, y esto implica que r divide a todos los coeficientes de $f(x)$ o bien a todos los coeficientes de $g(x)$.

En efecto, para probarlo basta considerar los polinomios de $\mathbb{Z}_r[x]$ que resultan de sustituir cada coeficiente por su clase de restos módulo r. Estos polinomios verifican la relación

$$\bar{f}(x)\bar{g}(x) = 0,$$

pero $\mathbb{Z}_r[x]$ es un dominio íntegro, si un producto de polinomios da 0, es que uno de los factores es 0, luego $\bar{f}(x) = 0$ o bien $\bar{g}(x) = 0$, lo que significa que r divide a los coeficientes de uno de los factores. Por lo tanto, podemos simplificar r de la ecuación $ac_p(x) = f(x)g(x)$, y tras un número finito de pasos tenemos que llegar a una ecuación de la forma

$$c_p(x) = f(x)g(x),$$

donde los factores están ahora en $\mathbb{Z}[x]$ y tienen grados $0 < u, v < p-1$. Como el coeficiente director de $c_p(x)$ vale 1, los coeficientes directores de $f(x)$ y $g(x)$ tienen que ser ± 1, y no perdemos generalidad si suponemos que ambos valen 1.

Ahora consideramos polinomios con coeficientes en $\mathbb{Z}_p[x]$. Notemos que

$$(x-1)c_p(x) = x^p - 1,$$

luego también

$$\bar{c}_p(x)(x-\bar{1}) = x^p - \bar{1} = (x-\bar{1})^p,$$

donde hemos usado el teorema 3.16, ya que que el anillo $\mathbb{Z}_p[x]$ tiene característica p. Por consiguiente,

$$(x-\bar{1})^{p-1} = \bar{c}_p(x) = \bar{f}(x)\bar{g}(x).$$

A su vez, esto implica que

$$\bar{f}(x) = \bar{c}(x-\bar{1})^u, \qquad \bar{g}(c) = \bar{d}(x-\bar{1})^v,$$

donde los exponentes u y v son precisamente los grados de $f(x)$ y $g(x)$ porque los coeficientes directores de estos polinomios valen 1, luego sus restos módulo p no son nulos, por lo que $\bar{f}(x)$ y $\bar{g}(x)$ tienen también grados u, v, respectivamente. Ahora, como $u, v > 0$, tenemos que $\bar{f}(\bar{1}) = \bar{0} = \bar{g}(\bar{1})$, pero esto equivale a que $p \mid f(1)$ y $p \mid g(1)$, luego $p^2 \mid f(1)g(1) = c_p(1) = p$ y tenemos una contradicción. ∎

En particular, ahora sabemos que hay polinomios irreducibles en $\mathbb{Q}[x]$ de grado arbitrariamente grande.

3.7 Restos potenciales

Vamos a investigar ahora cuándo un número entero tiene raíz n-sima módulo otro entero m:

Definición 3.35 Un entero a es un *resto potencial n-simo* módulo un entero m si existe un entero b tal que $a \equiv b^n \pmod{m}$. Los restos potenciales segundos, terceros y cuartos se llaman, respectivamente, restos *cuadráticos, cúbicos* y *bicuadráticos.*

La tabla siguiente muestra la situación para $m = 11$ y varios exponentes n. Vemos que cada potencia n-sima se repite un número d de veces, que es distinto para cada n, y que hemos recogido en la última columna. Tal vez el lector pueda conjeturar a partir de aquí (o a partir de otros ejemplos, preferentemente para otros módulos primos) la relación que existe entre n y d (y m). Por ejemplo, un hecho obvio es que, si las 10 potencias tienen que coincidir en grupos de d, necesariamente $d \mid 10$, de modo que exactamente $10/d$ de las 10 clases de restos (no nulas) módulo 11 son potencias n-simas.

x	1	2	3	4	5	6	7	8	9	10	1
x^2	1	4	9	5	3	3	5	9	4	1	2
x^3	1	8	5	9	4	7	2	6	3	10	1
x^4	1	5	4	3	9	9	3	4	5	1	2
x^5	1	10	1	1	1	10	10	10	1	10	5
x^6	1	9	3	4	5	5	4	3	9	1	2
x^7	1	7	9	5	3	8	6	2	4	10	1
x^8	1	3	5	9	4	4	9	5	3	1	2
x^9	1	6	4	3	9	2	8	7	5	10	1
x^{10}	1	1	1	1	1	1	1	1	1	1	10
x^{11}	1	2	3	4	5	6	7	8	9	10	1
x^{12}	1	4	9	5	3	3	5	9	4	1	2
x^{13}	1	8	5	9	4	7	2	6	3	10	1
x^{14}	1	5	4	3	9	9	3	4	5	1	2
x^{15}	1	10	1	1	1	10	10	10	1	10	5

En realidad el patrón que muestra la tabla anterior es consecuencia de un resultado general válido para grupos cíclicos finitos cualesquiera, y en particular es aplicable a los grupos de clases de restos U_m siempre que sean cíclicos, es decir, siempre que exista una raíz primitiva módulo m.

En este contexto general, consideremos un grupo cíclico G generado por un elemento g de orden m. Así, los elementos de G son $1, g, g^2, \ldots, g^{m-1}$. Nos planteamos cuántas soluciones tiene en G la ecuación $x^n = a$. Podemos expresar $a = g^b$ y $x = g^y$, con lo que el problema equivale a contar las soluciones y de la ecuación $g^{ny} = g^b$, con $0 \le y < m$.

Equivalentemente, la ecuación es $g^{ny-b} = 1$, y esto equivale a que $m \mid ny - b$, es decir, que buscamos el número de soluciones y de la congruencia

$$ny \equiv b \pmod{m}, \qquad 0 \le y < m.$$

Si $d = (n, m)$, para que esta congruencia tenga solución, es decir, para que exista un k tal que $ny = b + km$, es necesario que $d \mid b$, y en tal caso las

soluciones cumplen $(n/d)y = (b/d) + k(m/d)$, luego buscamos las soluciones de la congruencia

$$\frac{n}{d}y \equiv \frac{b}{d} \text{ (mód } \frac{m}{d}\text{)}, \qquad 0 \le y < m.$$

Como $(n/d, m/d) = 1$, esta congruencia tiene una solución única módulo m/d, luego ahora podemos afirmar, por una parte, que la ecuación tiene solución si y sólo si $d \mid b$ y, por otra parte, que tiene exactamente d soluciones, pues si $0 \le y_0 < m/d$ es la única sólución menor que m/d, también son soluciones los números

$$y = \quad y_0, \quad y_0 + \frac{m}{d}, \quad y_0 + 2\,\frac{m}{d}, \quad \dots \quad y_0 + (d-1)\frac{m}{d},$$

pues todos ellos cumplen $0 \le y < m$.

Al aplicar esto al grupo $G = U_{11}$, que tiene $m = 10$ elementos, concluimos que el valor d mostrado en la tabla anterior para cada exponente n es precisamente $d = (n, 10)$. Sin embargo, el análisis anterior nos permite expresar de una forma más simple (independiente del generador g) la condición $d \mid b$ que hemos visto que es necesaria y suficiente para que un elemento de G tenga raíz n-sima:

Teorema 3.36 *Sea G un grupo cíclico finito de orden m, sea $a \in G$, sea n un número entero y sea $d = (m, n)$. Entonces, la ecuación $x^n = a$ tiene solución en G si y sólo si $a^{m/d} = 1$, y en tal caso tiene exactamente d soluciones distintas.*

Demostración: Teniendo en cuenta la discusión precedente, en la que hemos expresado $a = g^b$, sólo queda demostrar que $d \mid b$ si y sólo si $a^{m/d} = 1$.

En efecto, si $d \mid b$ entonces $a^{m/d} = (g^b)^{m/d} = (g^m)^{n/d} = 1^{n/d} = 1$. Recíprocamente, si $a^{m/d} = 1$, esto equivale a que $g^{bm/d} = 1$, lo que a su vez equivale a que $m \mid bm/d$, es decir, a que exista un k tal que $km = bm/d$, o también $dk = b$, luego $d \mid b$. ∎

Si particularizamos el teorema anterior al caso de los grupos de unidades U_m tenemos el enunciado siguiente:

Teorema 3.37 *Sea m un número natural tal que exista una raíz primitiva módulo m, sea n un número entero y sea $d = (n, \phi(m))$. Si $(a, m) = 1$, entonces a es un resto potencial n-simo módulo m si y sólo si $a^{\phi(m)/d} \equiv 1$ (mód m), y en tal caso a tiene exactamente d raíces n-simas distintas módulo m. En particular, hay exactamente $\phi(m)/d$ restos potenciales n-simos en U_m.*

De aquí obtenemos un criterio sencillo debido a Euler que determina para qué primos impares p se cumple que -1 es un resto cuadrático módulo p. En este caso $d = (2, p-1) = 2$, luego el teorema anterior nos da que esto sucede cuando $(-1)^{(p-1)/2} \equiv 1$ (mód p), pero es obvio que esto sucede exactamente cuando el exponente es par, es decir, cuando $4 \mid p - 1$. Por consiguiente:

Teorema 3.38 (Euler) *Si p es un primo impar, entonces -1 es un resto cuadrático módulo p si y sólo si $p \equiv 1$ (mód 4).*

Ejercicio: Formular una conjetura a partir de ejemplos concretos sobre cuándo 2 es un resto cuadrático módulo un primo impar p.

Veamos otro ejemplo:

Teorema 3.39 *Si $p > 3$ es primo, entonces -3 es un resto cuadrático módulo p si y sólo si $p \equiv 1$ (mód 3).*

DEMOSTRACIÓN: Observamos que $x^3 - 1(x-1)(x^2+x+1)$, y el segundo polinomio tiene discriminante $\Delta = -3$, luego el teorema 3.17 nos da que -3 es un resto cuadrático módulo p si y sólo si dicho polinomio tiene dos raíces en $\mathbb{Z}_p$, si y sólo si $x^3 - 1$ tiene tres raíces en $\mathbb{Z}_p$ y, por el teorema 3.37 (aplicado a $a = 1$) esto equivale a que $(3, p-1) = 3$, es decir, a que $3 \mid p-1$. ∎

Ejercicio: Formular una conjetura a partir de ejemplos concretos sobre cuándo 3 es un resto cuadrático módulo un primo $p > 3$.

Capítulo IV

Los enteros de Gauss

4.1 Sumas de dos cuadrados I

Consideremos el problema siguiente ¿qué números naturales pueden expresarse en la forma $x^2 + y^2$? Aquí tenemos las sumas de dos cuadrados hasta 100:

0	1	2	4	5	8	9	10	13	16	17
18	20	25	26	29	32	34	36	37	40	41
45	49	50	52	53	58	61	64	65	68	72
73	74	80	81	82	85	89	90	97	98	100

Obviamente, en la tabla encontramos todos los cuadrados. Más aún, una observación elemental es que si $n = x^2 + y^2$ es suma de dos cuadrados, entonces $k^2n = (kx)^2 + (ky)^2$ también lo es, luego podemos separar de la tabla aquellos números cuya presencia se explica por ser de la forma k^2n con n ya presente en la tabla. Eliminamos así:

$4 = 2^2 \cdot 1$	$8 = 2^2 \cdot 2$	$9 = 3^2 \cdot 1$	$16 = 4^2 \cdot 1$	$18 = 3^2 \cdot 2$	$20 = 2^2 \cdot 5$
$25 = 5^2 \cdot 1$	$32 = 4^2 \cdot 2$	$36 = 6^2 \cdot 1$	$40 = 2^2 \cdot 10$	$45 = 3^2 \cdot 5$	$49 = 7^2 \cdot 1$
$50 = 5^2 \cdot 2$	$52 = 2^2 \cdot 13$	$64 = 8^2 \cdot 1$	$68 = 2^2 \cdot 17$	$72 = 6^2 \cdot 2$	$80 = 4^2 \cdot 5$
$81 = 9^2 \cdot 1$	$90 = 3^2 \cdot 10$	$98 = 7^2 \cdot 2$	$100 = 10^2 \cdot 1$		

y nos quedan:

0	1	**2**	**5**	$2 \cdot 5$	**13**	**17**	$2 \cdot 13$	**29**	$2 \cdot 17$	**37**
41	**53**	$2 \cdot 29$	**61**	$5 \cdot 13$	**73**	$2 \cdot 37$	$2 \cdot 41$	$5 \cdot 17$	**89**	**97**

Hemos descompuesto los números restantes en factores primos porque así se ponen de manifiesto dos hechos:

1. Hemos probado que si n es suma de dos cuadrados, también lo es k^2n, pero en principio podría ocurrir que k^2n fuera suma de dos cuadrados sin que n lo sea. Sin embargo, en la tabla no ha aparecido ningún número en estas circunstancias, pues ninguno de los números que han quedado al eliminar los de la forma k^2n con n en la tabla es de la forma k^2n.

Esto induce a conjeturar que un número k^2n es suma de dos cuadrados si y sólo si n es suma de dos cuadrados.

2. Los factores primos de los números que han quedado en la tabla están también en la tabla.

Si estos hechos no son casuales, podemos formular una conjetura muy completa sobre la situación. Observemos que todo número entero no nulo m puede expresarse de forma única como $m = k^2n$, donde n es libre de cuadrados. Se dice que n es la *parte libre de cuadrados* de m, y está formada por los primos que dividen a m con exponente impar.

En estos términos, es razonable conjeturar:

1. Un número natural no nulo es suma de dos cuadrados si y sólo si lo es su parte libre de cuadrados.

2. Un número natural no nulo libre de cuadrados es suma de dos cuadrados si y sólo si sus divisores primos son suma de dos cuadrados.

Notemos que 2. no puede ser cierto para números cualesquiera porque todo número de la forma $k^2 \cdot 5$ es suma de dos cuadrados, y k puede ser divisible por primos cualesquiera.

Por último nos fijamos en los primos que son suma de dos cuadrados:

$$2,\ 5,\ 13,\ 17,\ 29,\ 37,\ 41,\ 53,\ 61,\ 73,\ 89,\ 97,\ \ldots$$

¿Identifica el lector esta sucesión? Una conjetura que (de ser cierta) resuelve completamente el problema es:

> *Un número natural no nulo es suma de dos cuadrados si y sólo si los primos que lo dividen con exponente impar son el* 2 *o bien primos congruentes con* 1 *módulo* 4.

Vamos a ver que la conjetura es exacta. Consideremos en primer lugar la fórmula

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2.$$

Esta igualdad nos garantiza que un producto de números expresables como suma de dos cuadrados es también expresable como suma de dos cuadrados. El recíproco no es cierto, pero algo podemos probar:

Veamos que si $a = pb$ donde p es primo y tanto a como p son suma de dos cuadrados, entonces b también lo es.

Para ello observemos que si $p = u^2 + v^2$ y $a = r^2 + s^2$, entonces

$$\begin{aligned}(us - rv)(us + rv) &= u^2s^2 - r^2v^2 = u^2s^2 + u^2r^2 - u^2r^2 - r^2v^2\\ &= u^2(s^2 + r^2) - r^2(u^2 + v^2) = u^2a - r^2p,\end{aligned}$$

luego $p \mid (us - rv)(us + rv)$ y por tanto $p \mid (us - rv)$ o bien $p \mid (us + rv)$.

Si $p \mid (us + rv)$, entonces, como $p^2 \mid ap$ y

$$ap = (r^2 + s^2)(u^2 + v^2) = (ru - sv)^2 + (rv + us)^2,$$

resulta que $p^2 \mid (ru - sv)^2$, luego

$$b = \frac{a}{p} = \frac{(ru - sv)^2}{p^2} + \frac{(rv + us)^2}{p^2}$$

es suma de cuadrados.

Si $p \mid (us - rv)$ razonamos igual con la fórmula

$$ap = (s^2 + r^2)(u^2 + v^2) = (su - rv)^2 + (sv + ur)^2.$$

Esto implica que si un número a, expresable como suma de dos cuadrados, es divisible por un número b que no lo es, entonces el cociente tiene un factor primo no expresable como suma de dos cuadrados, pues tendríamos $a = bc$ y, si todos los factores primos de c fueran expresables como suma de dos cuadrados, una aplicación repetida del resultado anterior nos daría que b es también expresable como suma de dos cuadrados.

Ahora viene el resultado fundamental: si p es un primo tal que $(a, b) = 1$ y $p \mid (a^2 + b^2)$, entonces p es suma de dos cuadrados. Notemos que si r es libre de cuadrados y $r = a^2 + b^2$, entonces $(a, b) = 1$, pues si $d \mid a$ y $d \mid b$, entonces $d^2 \mid r$. Por lo tanto un número libre de cuadrados es suma de dos cuadrados si y sólo si sus factores primos lo son.

En efecto, sea $a = pm \pm c$, $b = pn \pm d$, donde $|c|$, $|d| \leq p/2$ (si el resto de la división resulta mayor que $p/2$ sumamos 1 al cociente y tomamos resto negativo).

Entonces $a^2 + b^2 = m^2p^2 \pm 2mpc + c^2 + n^2p^2 \pm 2npd + d^2 = Ap + (c^2 + d^2)$. En consecuencia $p \mid (c^2 + d^2)$, o sea, $c^2 + d^2 = py$, para cierto y. Como $(c, d) < p$, p no lo divide, luego $(c, d)^2 \mid y$.

Ahora dividimos la ecuación $c^2 + d^2 = py$ hasta obtener $e^2 + f^2 = pz$, donde $(e, f) = 1$ y $pz \leq c^2 + d^2 \leq (p/2)^2 + (p/2)^2 = p^2/2$, luego $z \leq p/2$.

Si p no fuera suma de dos cuadrados, por el resultado anterior z tiene un factor primo q que tampoco es expresable como suma de dos cuadrados. En particular $q \leq z < p$.

Hemos probado que si existen números p, a, b, tales que p es primo, $(a, b) = 1$ y p divide a $a^2 + b^2$, entonces existen números q, e, f, en las mismas condiciones y con $q < p$. Pero si existieran tales ternas de números debería haber una con p mínimo, y según lo visto es imposible.

Esto prueba la mayor parte de nuestra conjetura: Sea $n = u^2v$ donde v es libre de cuadrados y supongamos que n es suma de dos cuadrados,

$$n = a^2 + b^2 = (a, b)^2(c^2 + d^2)$$

con $(c, d) = 1$. Entonces $(a, b) \mid u$, luego $v \mid (c^2 + d^2)$. Por el resultado que acabamos de probar todo primo que divide a v (luego a $c^2 + d^2$) es suma de dos cuadrados, luego v es suma de dos cuadrados.

En resumen, tenemos probado que un número es suma de dos cuadrados si y sólo si lo es su parte libre de cuadrados, si y sólo si los primos que dividen a su parte libre de cuadrados son suma de dos cuadrados. Sólo falta probar que los únicos primos impares expresables como suma de dos cuadrados son exactamente los congruentes con 1 módulo 4.

La condición es obviamente necesaria: si $p = x^2 + y^2$, entonces x e y deben tener paridades opuestas, luego uno de los cuadrados x^2 o y^2 tiene que ser congruente con 0 módulo 2 y el otro con 1, luego la suma es $p \equiv 1 \pmod 4$.

Recíprocamente, supongamos que $p = 4n + 1$. Por el teorema de Fermat sabemos que todo número a entre 1 y $p-1$ cumple $a^{p-1} \equiv 1 \pmod p$, es decir, $a^{4n} \equiv 1 \pmod p$, luego $(a+1)^{4n} - a^{4n} \equiv 0 \pmod p$, es decir, $p \mid (a+1)^{4n} - a^{4n}$ para $1 \leq a \leq 4n-1$. Por otra parte

$$(a+1)^{4n} - a^{4n} = ((a+1)^{2n} + a^{2n})((a+1)^{2n} - a^{2n}),$$

luego o bien $p \mid (a+1)^{2n} + a^{2n}$ o bien $p \mid (a+1)^{2n} - a^{2n}$.

Si $p \mid (a+1)^{2n} + a^{2n}$ para algún número a, entonces, como $(a, a+1) = 1$, sabemos que p es suma de dos cuadrados. Veamos que no es posible que p no divida a ninguno de estos números, es decir, que no es posible que se cumpla $p \mid (a+1)^{2n} - a^{2n}$ para todo número a entre 1 y $4n-1$. En tal caso p divide a las $4n-2$ diferencias de las potencias $2n$–simas de los $4n-2$ primeros números enteros, luego también a las $4n-3$ diferencias de sus diferencias, etc. y así debería dividir a las $2n$ diferencias de orden $2n$, que, según hemos visto en el ejemplo de la página 30, valen $(2n)!$, pero como p es primo, resulta que p divide a un $m \leq 2n$, lo cual es imposible ya que $p = 4n+1$.

Lo que acabamos de ver es un ejemplo de demostración típica de finales del siglo XVIII. Platón diría que es una demostración que maneja las "sombras de las ideas" en lugar de las ideas en sí. En la sección siguiente vamos a introducir las "ideas" necesarias para dar una demostración mucho más conceptual, que descansa sobre argumentos claros y no sobre meros cálculos que "funcionan".

4.2 Enteros de Gauss

En el capítulo anterior hemos visto que el concepto de congruencia está definido sobre anillos arbitrarios. En particular, el teorema 3.3 nos da que las clases de restos módulo cualquier elemento de cualquier anillo forman a su vez un anillo (que será un dominio si partimos de un dominio y el módulo no es una unidad).

Vamos a aprovechar esta generalidad para considerar el anillo de clases de restos en el anillo de polinomios $A = \mathbb{Q}[x]$ módulo el polinomio $p(x) = x^2 + 1$.

Como $p(x)$ no tiene raíces en $\mathbb{Q}$ (ya que -1 no es el cuadrado de ningún número racional) resulta que $p(x)$ es irreducible (que es lo mismo que primo). El mismo argumento que hemos empleado en $\mathbb{Z}$ nos da lo siguiente:

Teorema 4.1 *Si A es un dominio euclídeo y p es un primo en A, entonces el anillo de clases de restos A_p es un cuerpo.*

DEMOSTRACIÓN: Sabemos que A_p es un dominio. Basta probar que todo elemento no nulo tiene inverso. Tomemos una clase $\bar{a} \neq \bar{0}$ en A_p. Esto significa que $p \nmid a$, luego $(a, p) = 1$. Por la relación de Bezout 2.20, existen elementos u, v en A tales que $ua + vb = 1$, y tomando clases módulo p resulta que $\bar{u}\bar{a} = \bar{1}$, luego $\bar{a}$ tiene como inverso a la clase $\bar{u}$. ∎

Así pues, tenemos que el anillo de clases de restos $k = \mathbb{Q}[x]_{x^2+1}$ es un cuerpo. Vamos a analizarlo con detalle.

En principio, un elemento arbitrario de k es de la forma $\overline{q(x)}$, donde $q(x)$ es un polinomio arbitrario en $\mathbb{Q}[x]$. Ahora bien, podemos dividir

$$q(x) = (x^2 + 1)c(x) + a + bx,$$

para ciertos números racionales a, b (ya que el resto tiene que ser un polinomio de grado menor que 2). Al tomar clases módulo $x^2 + 1$ queda que

$$\overline{q(x)} = \bar{a} + \bar{b}\bar{x}.$$

Con esto casi hemos probado lo siguiente:

> *Todo elemento de k se expresa de forma única en la forma $\bar{a} + \bar{b}\bar{x}$, donde a y b son números racionales.*

Sólo falta probar la unicidad, pero si $\bar{a} + \bar{b}\bar{x} = \bar{c} + \bar{d}\bar{x}$, entonces

$$\overline{a - c + (b - d)x} = \bar{0},$$

lo cual equivale a que $x^2 + 1 \mid a - c + (b - d)x$, pero un polinomio de grado 2 sólo puede dividir a un polinomio de grado menor que 2 si éste es nulo, luego $a - c + (b - d)x = 0$, lo cual significa que $a = c$ y $b = d$.

En particular, la aplicación $\mathbb{Q} \longrightarrow k$ dada por $a \mapsto \bar{a}$ hace corresponder números racionales distintos con elementos de k distintos. Además, tenemos que

$$\overline{a + b} = \bar{a} + \bar{b}, \qquad \overline{ab} = \bar{a}\bar{b},$$

y esto significa que podemos identificar cada número racional a con la clase $\bar{a}$ de k y así considerar que $\mathbb{Q}$ está contenido en k (en otros términos, hemos probado que k es un cuerpo de característica 0).

Con esta identificación podemos decir que los elementos de k son de la forma $a + b\bar{x}$, donde a, b son números racionales. Por último observamos que, trivialmente, $\overline{x^2 + 1} = \bar{0}$, luego $\bar{x}^2 = -1$. Si convenimos en llamar $i = \bar{x}$, resulta que los elementos de k son de la forma $a + bi$, donde a y b son números racionales y además $i^2 = -1$.

Definición 4.2 Llamaremos $\mathbb{Q}(i)$ al cuerpo de clases de restos de $\mathbb{Q}[x]$ módulo el polinomio x^2+1. Llamando $i = \bar{x}$ e identificando cada número racional a con la clase $\bar{a}$, tenemos que los elementos de $\mathbb{Q}(i)$ se expresan de forma única como $a + bi$ con a y b números racionales, y además se cumple que $i^2 = -1$.

Teniendo en cuenta además que la suma y el producto de $\mathbb{Q}(i)$ se restringe a las operaciones usuales sobre los números racionales, estos hechos determinan completamente las operaciones en $\mathbb{Q}(i)$. En efecto:

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$

mientras que para multiplicar dos de estos números sólo tenemos que operar:

$$(a + bi)(c + di) = ac + bdi^2 + adi + bci = (ac - bd) + (ad + bc)i.$$

Nota Podríamos haber introducido $\mathbb{Q}(i)$ más directamente, definiendo la suma y el producto con las fórmulas explícitas que acabamos de dar, pero entonces tendríamos que demostrar que tales operaciones cumplen todas las propiedades de la definición de cuerpo, lo cual es un tanto laborioso. La construcción como anillo de clases de restos nos asegura directamente que $\mathbb{Q}(i)$ es un cuerpo.[1] Vamos a apoyarnos en esta construcción para obtener un último hecho sobre $\mathbb{Q}(i)$, pero, una vez lo tengamos, podremos olvidarnos de que los elementos de $\mathbb{Q}(i)$ son clases de restos de polinomios y considerar únicamente que son números de la forma $a + bi$. ∎

El último hecho básico que tenemos que comprobar es que si definimos el *conjugado* de un número $a + bi$ como $\overline{a + bi} = a - bi$ (no confundir la barra con la notación que usamos para representar las clases de restos), se cumple que

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \qquad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2.$$

Esto puede probarse mediante un cálculo explícito (que el lector puede llevar a cabo como ejercicio) o también a partir de la construcción que hemos hecho de $\mathbb{Q}(i)$. Para ello, consideramos la aplicación

$$c : \mathbb{Q}[x] \longrightarrow \mathbb{Q}(i)$$

dada por $c(p(x)) = p(-i)$. En general, la evaluación de polinomios en un elemento dado cumple las propiedades siguientes:

$$c(p(x) + q(x)) = p(-i) + q(-i) = c(p(x)) + c(q(x)),$$

$$c(p(x)q(x)) = p(-i)q(-i) = c(p(x))x(q(x)).$$

[1] El lector familiarizado con los números complejos habrá advertido que podríamos haber definido $\mathbb{Q}(i)$ como el subcuerpo de $\mathbb{C}$ formado por los números complejos $a + bi$ en los que a y b son números racionales. Si en la construcción que hemos hecho cambiamos $\mathbb{Q}$ por $\mathbb{R}$, obtenemos una construcción del cuerpo $\mathbb{C}$ de los números complejos sin ninguna otra variación.

Además, $c(x^2+1) = (-i)^2 + 1 = 0$, luego si $p(x) \equiv q(x)$ (mód x^2+1), esto significa que $p(x) - q(x) = (x^2+1)h(x)$ y, evaluando en $-i$, queda que $c(p(x)) - c(q(x)) = 0$, luego $c(p(x)) = c(q(x))$. Por consiguiente, c induce una aplicación

$$\bar{c} : \mathbb{Q}(i) \longrightarrow \mathbb{Q}(i)$$

dada por $\bar{c}([p(x)]) = c(p(x))$, que está bien definida porque hemos visto que no importa qué polinomio $p(x)$ tomamos como representante de la clase de restos. Es inmediato que $\bar{c}$ cumple también

$$\bar{c}(z_1 + z_2) = \bar{c}(z_1) + \bar{c}(z_2), \qquad \bar{c}(z_1 z_2) = \bar{c}(z_1)\bar{c}(z_2).$$

Ahora bien, $\bar{c}$ no es sino la conjugación que hemos definido antes, pues

$$\bar{c}(a + bi) = \bar{c}([a + bx]) = a - bi.$$

Llegados a este punto, el lector puede, si lo desea, olvidar toda la construcción que hemos hecho y quedarse únicamente con las propiedades recogidas en en resumen 4.1. Hemos demostrado ya las cuatro primeras, y las restantes se deducen fácilmente de ellas.

En primer lugar tenemos la definición de la norma de un número $z = a + bi$ como

$$\mathrm{N}(z) = z\bar{z} = (a + bi)(a - bi) = a^2 + b^2.$$

Observemos que si $z \neq 0$, entonces $\mathrm{N}(z) \neq 0$. Obviamente

$$\mathrm{N}(z_1 z_2) = z_1 z_2 \bar{z}_1 \bar{z}_2 = \mathrm{N}(z_1)\,\mathrm{N}(z_2).$$

Además, ahora podemos obtener una expresión explícita para el inverso de un elemento (no nulo) de $\mathbb{Q}(i)$. Es obvio que

$$z \cdot \frac{\bar{z}}{\mathrm{N}(z)} = \frac{\mathrm{N}(z)}{\mathrm{N}(z)} = 1,$$

luego

$$z^{-1} = \frac{\bar{z}}{\mathrm{N}(z)}.$$

Por ejemplo, para calcular el inverso de $z = 3 + 2i$ calculamos su norma $\mathrm{N}(z) = 3^2 + 2^2 = 13$ y entonces

$$z^{-1} = \frac{3 - 2i}{13} = \frac{3}{13} - \frac{2}{13}i.$$

Más en general, para dividir dos elementos de $\mathbb{Q}(i)$ basta multiplicar por el conjugado del denominador:

$$\frac{2 + i}{3 - 4i} = \frac{(2 + i)(3 + 4i)}{(3 - 4i)(3 + 4i)} = \frac{2 + 11i}{25} = \frac{2}{25} + \frac{11}{25}i.$$

Nota La notación $\mathbb{Q}(i)$ sugiere que este cuerpo se obtiene "adjuntándole" a $\mathbb{Q}$ una raíz cuadrada de -1. Se trata del menor cuerpo que podemos construir que

Resumen 4.1: El cuerpo $\mathbb{Q}(i)$

- $\mathbb{Q}(i)$ es un cuerpo cuyos elementos se expresan de forma única como $a+bi$, donde a y b son números racionales e $i^2 = -1$.
- La suma y el producto en $\mathbb{Q}(i)$ vienen dadas por

$$(a+bi)+(c+di) = (a+c)+(b+d)i, \quad (a+bi)(c+di) = (ac-bd)+(ad+bc)i.$$

- El *conjugado* de un elemento de $\mathbb{Q}(i)$ es $\overline{a+bi} = a - bi$.
- La conjugación cumple las propiedades

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \qquad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2.$$

- La *norma* $\mathrm{N} : \mathbb{Q}(i) \longrightarrow \mathbb{Q}$ se define como

$$\mathrm{N}(z) = z\bar{z}, \qquad \text{o equivalentemente.} \qquad \mathrm{N}(a+bi) = a^2 + b^2.$$

- Se cumple que $\mathrm{N}(z_1 z_1) = \mathrm{N}(z_1)\,\mathrm{N}(z_2)$.
- Si $z \neq 0$ es un elemento de $\mathbb{Q}(i)$, entonces

$$z^{-1} = \frac{\bar{z}}{\mathrm{N}(z)}.$$

contenga a $\mathbb{Q}$ de modo que -1 tenga raíz cuadrada. Más concretamente, es fácil convencerse de que cualquier cuerpo que contenga a $\mathbb{Q}$ y en el que -1 tenga una raíz cuadrada tendrá un subcuerpo que podemos identificar con $\mathbb{Q}(i)$. ∎

En realidad no estamos tan interesados en $\mathbb{Q}(i)$ como en un anillo contenido en él:

Definición 4.3 El anillo de los *enteros de Gauss* es el anillo $\mathbb{Z}[i]$ formado por los elementos de $\mathbb{Q}(i)$ de la forma $a + bi$, donde a y b son números enteros.

Es inmediato que al sumar, restar o multiplicar enteros de Gauss obtenemos un entero de Gauss, lo cual se traduce en que, ciertamente $\mathbb{Z}[i]$ es un anillo. De hecho es un dominio íntegro, ya que está contenido en un cuerpo, luego no puede tener divisores de 0.

Lo que ya no es cierto es que al dividir enteros de Gauss obtengamos un entero de Gauss. Por el contrario, es fácil ver que el cuerpo de cocientes de $\mathbb{Z}[i]$ es $\mathbb{Q}(i)$. Por una parte, es inmediato que al formar fracciones con elementos de $\mathbb{Z}[i]$ obtenemos elementos de $\mathbb{Q}(i)$ y, por otra parte, todo elemento de $\mathbb{Q}(i)$ es cociente de enteros de Gauss. Más aún, es un cociente de un entero de Gauss

entre un entero ordinario. Por ejemplo:

$$\frac{3}{5}+\frac{2}{7}i=\frac{21}{35}+\frac{10}{35}i=\frac{21+10i}{35},$$

y lo mismo puede hacerse con cualquier otro elemento de $\mathbb{Q}(i)$.

Para distinguir a los enteros de Gauss de los enteros usuales, a éstos se les suele llamar en este contexto *enteros racionales*, de modo que los enteros de Gauss son los "enteros" del cuerpo $\mathbb{Q}(i)$, mientras que los enteros racionales son los enteros del cuerpo $\mathbb{Q}$ de los números racionales.

Un hecho elemental que va a ser muy relevante es que, mientras la norma de un elemento de $\mathbb{Q}(i)$ es un número racional, al considerar enteros de Gauss tenemos que $\mathrm{N}:\mathbb{Z}[i]\longrightarrow\mathbb{N}$.

El hecho fundamental sobre enteros de Gauss es el teorema siguiente:

Teorema 4.4 *$\mathbb{Z}[i]$ es un dominio euclídeo con la norma $\mathrm{N}:\mathbb{Z}[i]\longrightarrow\mathbb{N}$.*

Demostración: Teniendo en cuenta que las normas de los enteros de Gauss son números naturales, es evidente que si α y β son enteros de Gauss no nulos, entonces $\mathrm{N}(\alpha)\le\mathrm{N}(\alpha)\,\mathrm{N}(\beta)=\mathrm{N}(\alpha\beta)$.

Tomemos ahora dos enteros de Gauss Δ y δ con $\delta\neq 0$. Consideramos el cociente $\Delta/\delta=r+si$, donde r y s son números racionales. Llamamos x e y a los enteros racionales más cercanos a r y s, respectivamente, de modo que $|r-x|$, $|s-y|\le 1/2$, con lo que, llamando $\gamma=x+yi$, $\epsilon=\Delta-\delta\gamma$, tenemos que

$$\mathrm{N}(\epsilon/\delta)=\mathrm{N}(\Delta/\delta-\gamma)=(r-x)^2+(s-y)^2\le 1/4+1/4=1/2<1,$$

luego $\mathrm{N}(\epsilon)<\mathrm{N}(\delta)$. ∎

Observemos que la demostración nos dice cómo obtener explícitamente el cociente y el resto de una división: para obtener el cociente realizamos la división en $\mathbb{Q}(i)$ y redondeamos las coordenadas del resultado hacia los enteros más próximos, y una vez conocido el cociente ya podemos calcular el resto.

Ya sabemos la trascendencia que tiene el teorema anterior: ahora sabemos que los enteros de Gauss son un dominio de factorización única en el que podemos manejar los conceptos aritméticos igual que en $\mathbb{Z}$. Por ejemplo, la tabla siguiente muestra una aplicación del algoritmo de Euclides para calcular el máximo común divisor de los enteros de Gauss $36+37i$ y $-16+37i$. Indicamos también las normas de los restos que vamos calculando:

c	r	u	v	N
	$36+37i$	1	0	2 665
$-i$	$-16+37i$	0	1	1 625
$2+i$	$-1+21i$	1	i	442
$-1+2i$	$7-4i$	$-2-i$	$2-2i$	65
$-2-i$	$-2+3i$	$-3+3i$	$-2-5i$	13
	0			

Por ejemplo, la segunda división que hemos hecho ha sido

$$\frac{-16+37i}{-1+21i} = \frac{793+299i}{442} = 1.79 + 0.68i \approx 2 + i,$$

lo que nos da un resto de $-16+37i-(-1+21i)(2+i) = 7-4i$.

La conclusión es que

$$(36+37i, -16+37i) = -2+3i,$$

con la relación de Bezout:

$$-2+3i = (-3+3i)(36+37i) + (-2-5i)(-16+37i).$$

Unidades Para "orientarnos" en $\mathbb{Z}[i]$ lo primero que necesitamos tener claro es cuáles son las unidades de este anillo. Observemos que las unidades son necesariamente los enteros de norma 1. En efecto, si α es un entero de Gauss y $\mathrm{N}(\alpha) = \alpha\bar{\alpha} = 1$, entonces el conjugado $\bar{\alpha}$ también es un entero de Gauss y vemos así que es el inverso de α, luego α es una unidad. Recíprocamente, si α es una unidad, tenemos que $\mathrm{N}(\alpha)\,\mathrm{N}(\alpha^{-1}) = \mathrm{N}(\alpha\alpha^{-1}) = \mathrm{N}(1) = 1$ y, como las normas son números naturales, tiene que ser $\mathrm{N}(\alpha) = 1$.

Por consiguiente, $\alpha = x + yi$ es una unidad si y sólo si $x^2 + y^2 = 1$ y, puesto que x, y son números enteros, es obvio que esto sólo puede suceder si $(x, y) = (\pm 1, 0)$, $(0, \pm 1)$. En conclusión:

Las unidades de $\mathbb{Z}[i]$ son 1, -1, i, $-i$.

Así pues, mientras que en $\mathbb{Z}$ los enteros no nulos se distribuyen en parejas de asociados, en $\mathbb{Z}[i]$ cada entero no nulo tiene cuatro asociados (incluyéndolo a él). Por ejemplo:

$$3+5i, \qquad -5+3i, \qquad -3-5i, \qquad 5-3i$$

son cuatro enteros de Gauss asociados entre sí. Los tres últimos resultan de multiplicar el primero por i, -1 y $-i$, respectivamente.

Nota No es cierto que todo elemento de $\mathbb{Q}(i)$ de norma 1 sea una unidad. Por ejemplo,

$$\frac{3}{5} + \frac{4}{5}i$$

tiene norma 1, pero no es una unidad porque no es un entero de Gauss.

Primos Vamos ahora a identificar a los primos de Gauss. Hay una condición suficiente muy simple para que un entero de Gauss sea primo:

Todo entero de Gauss de norma prima es primo.

En efecto, si α es un entero de Gauss y $\mathrm{N}(\alpha) = p$ es un número primo, ciertamente α no es nulo ni una unidad, y si factoriza como $\alpha = \beta\gamma$, entonces, tomando normas, $p = \mathrm{N}(\beta)\,\mathrm{N}(\gamma)$, por lo que $\mathrm{N}(\beta) = 1$ o bien $\mathrm{N}(\gamma) = 1$, es decir, que β o γ es una unidad. Esto prueba que α es irreducible, que en un dominio de factorización única es lo mismo que ser primo.

Por ejemplo, el número $\alpha = -2+3i$ que hemos obtenido en la aplicación del algoritmo de Euclides tiene norma 13, luego es un primo de Gauss.

Sin embargo, no es necesariamente cierto que un primo de Gauss deba tener norma prima. Para entender la situación general, consideremos un primo de Gauss arbitrario π. En general, todo entero de Gauss α divide a su norma, pues $\mathrm{N}(\alpha) = \alpha\bar{\alpha}$, y $\bar{\alpha}$ también es un entero de Gauss. En nuestro caso, $\pi \mid \mathrm{N}(\pi)$. Si descomponemos $\mathrm{N}(\pi)$ en factores primos (en $\mathbb{Z}$), el hecho de que π sea primo implica que debe dividir a uno de ellos. Así pues:

Todo primo de Gauss divide a un único primo racional.

Nos falta probar la unicidad, pero es trivial, pues si $\pi \mid p$, entonces tenemos que $\mathrm{N}(\pi) \mid \mathrm{N}(p) = p^2$, y es imposible que $\mathrm{N}(\pi)$ divida a la vez a p^2 y a q^2, para dos primos racionales distintos p y q.

En otras palabras, para encontrar los primos de Gauss sólo tenemos que descomponer en factores primos (de Gauss) los primos racionales.

Por ejemplo, tenemos que

$$5 = (2+i)(2-i)$$

y, como los dos factores tienen norma 5, son primos de Gauss. Conviene introducir algo de vocabulario:

Definición 4.5 Si p es un primo racional, se dice que:

1. p se *ramifica* en $\mathbb{Z}[i]$ si se descompone como $p = \epsilon\pi^2$, donde ϵ es una unidad y π es primo de norma p.

2. p se *escinde* en $\mathbb{Z}[i]$ si se descompone en dos factores primos no asociados $p = \pi_1\pi_2$ (de norma p).

3. p se *conserva* (primo) en $\mathbb{Z}[i]$ si es primo en $\mathbb{Z}[i]$ (y tiene norma p^2).

Observemos que no hay más posibilidades: Si la descomposición de un primo p en factores primos de Gauss es $p = \pi_1 \cdots \pi_r$, tomando normas vemos que $p^2 = \mathrm{N}(\pi_1)\cdots\mathrm{N}(\pi_r)$ y, como las normas de los primos son números naturales mayores que 1, o bien hay un único factor de norma p^2, o bien hay dos factores de norma p. En el primer caso p se conserva y en el segundo p se ramifica o se escinde según si los dos factores son asociados o no.

Más precisamente, si $\pi = a+bi$ es un factor primo de p de norma p, entonces $p = \mathrm{N}(a+bi) = a^2+b^2$ es suma de dos cuadrados y, recíprocamente, si $p = a^2+b^2$ es suma de dos cuadrados, entonces

$$p = (a+bi)(a-bi)$$

es una descomposición de p en factores primos de norma p. Por lo tanto:

> *Un primo racional p se ramifica o se escinde en $\mathbb{Z}[i]$ si y sólo si es suma de dos cuadrados, y en tal caso $p = \pi\bar{\pi}$ se descompone en producto de dos factores primos conjugados (asociados o no).*

Por lo tanto, para cada primo p en particular es fácil determinar cuál es su situación. Por ejemplo, como $2 = 1^2 + 1^2$, tenemos que

$$2 = (1+i)(1-i) = -i(1+i)^2,$$

es decir, 2 se descompone en producto de dos primos de norma 2, pero resultan ser asociados, por lo que 2 se ramifica.

En cambio 3 no es suma de dos cuadrados, luego se conserva primo en $\mathbb{Z}[i]$, mientras que $5 = 2^2 + 1^2$, luego

$$5 = (2+i)(2-i)$$

y es fácil ver que $2-i$ no es ninguno de los cuatro asociados de $2+i$, luego 5 se escinde.

Nota En la sección anterior hemos demostrado que un primo p es suma de dos cuadrados si y sólo si $p = 2$ o bien $p \equiv 1 \pmod 4$. Luego podemos afirmar que p se ramifica o se escinde si y sólo si $p = 2$ o $p \equiv 1 \pmod 4$. Sin embargo, hemos llegado a esta conclusión simplemente porque ciertos cálculos muestran que es así. A continuación vamos a probar un teorema que nos dará condiciones claras y naturales que determinan cuándo un primo se ramifica, se escinde o se conserva en $\mathbb{Z}[i]$, las cuales nos permitirán llegar a la misma conclusión que en la sección anterior hemos obtenido mediante meros cálculos. ∎

Sea p un primo racional y sea π un primo de Gauss que lo divida. Podemos considerar entonces el cuerpo de clases de restos de $\mathbb{Z}[i]$ módulo π, al que representaremos por k_π. El teorema 4.1 nos asegura que realmente es un cuerpo.

Los elementos de k_π son clases de restos de la forma $\overline{a+bi} = \bar{a} + \bar{b}\bar{i}$, donde a y b son enteros racionales. En principio podría haber infinitas clases así, pero en realidad sólo hay un número finito de clases, ya que si a y a' son dos números racionales, se cumple que

$$a \equiv a' \pmod \pi \quad \text{si y sólo si } a \equiv a' \pmod p.$$

En efecto, si $\pi \mid a-a'$, entonces $p \mid \mathrm{N}(\pi) \mid \mathrm{N}(a-a') = (a-a')^2$, luego $p \mid a-a'$. El recíproco es trivial. Esto nos da dos consecuencias:

1. Podemos identificar cada clase de restos $\bar{a}$ de $\mathbb{Z}_p$ con la clase correspondiente en k_π, y así podemos ver a $\mathbb{Z}_p$ como un subcuerpo de k_π.

2. Según esta identificación, cada elemento de k_π se expresa en la forma $a+b\bar{i}$, con a, $b \in \mathbb{Z}_p$, luego k_π tiene a lo sumo p^2 elementos.

Consideremos por ejemplo $p = \pi = 3$, que es un primo de Gauss. Lo que acabamos de probar es que k_3 está formado por las clases de los nueve enteros siguientes:

$2i$	$1+2i$	$2+2i$
i	$1+i$	$2+i$
0	1	2

de modo que las clases de la última fila forman un subcuerpo que podemos identificar con $\mathbb{Z}_3$.

Ejercicio: Calcular las tablas de la suma y el producto del cuerpo k_3.

Ejercicio: Encontrar una raíz primitiva en k_3.

Ahora bien, no es necesariamente cierto que k_π tenga p^2 elementos, sino que vamos a probar que en general tiene $\mathrm{N}(\pi)$ elementos (es decir, p o p^2).

Si $\mathrm{N}(\pi) = p^2$, es decir, si el primo racional p se conserva primo en $\mathbb{Z}[i]$, se cumple que $p \mid a+bi$ si y sólo si existe otro entero de Gauss $u+vi$ de modo que

$$a + bi = p(u + vi) = pu + pvi,$$

y esto equivale a que $a = pu$ y $b = pv$, o también a que $p \mid a$ y $p \mid b$. Más en general:

$$a + bi \equiv c + di \pmod{p} \quad \text{si y sólo si} \quad a \equiv c \pmod{p},\ b \equiv d \pmod{p}.$$

Esto se traduce en que k_π tiene en este caso p^2 elementos, porque los enteros $a + bi$, con $0 \le a < p$, $0 \le b < p$ son no congruentes dos a dos módulo p.

Así pues, cuando $p = \pi$ se conserva primo, se cumple que k_π tiene $\mathrm{N}(\pi) = p^2$ elementos, como habíamos afirmado.

Consideremos ahora el caso en que $\mathrm{N}(\pi) = p$, es decir, el caso en que p se ramifica o se escinde. Entonces $\pi = x + yi$, con $x^2 + y^2 = p$.

Por ejemplo, si $p = 5$ y $\pi = 2 + i$, sabemos que todo elemento de k_π es la clase de uno de los 25 enteros siguientes:

$4i$	$1+4i$	$2+4i$	$3+4i$	$4+4i$
$3i$	$1+3i$	$2+3i$	$3+3i$	$4+3i$
$2i$	$1+2i$	$2+2i$	$3+2i$	$4+2i$
i	$1+i$	$2+i$	$3+i$	$4+i$
0	1	2	3	4

y sabemos que las de la última fila forman un cuerpo que podemos identificar con $\mathbb{Z}_5$, pero ahora no es cierto que las 25 clases de restos sean distintas entre sí. De hecho, $k_\pi = \mathbb{Z}_5$.

El argumento que lo prueba vale en general: como $x^2 + y^2 = p$ y el primo p no puede ser un cuadrado, vemos que $0 < y < p$, lo que implica que la clase de restos $\bar{y}$ en $\mathbb{Z}_p$ no es nula, luego tiene una clase inversa, es decir, existe un entero racional z tal que $\bar{z}\bar{y} = \bar{1}$.

Por otra parte, $\bar{x} + \bar{y}\bar{i} = \bar{\pi} = \bar{0}$, luego podemos despejar $\bar{y}\bar{i} = -\bar{x}$ y a su vez $\bar{\imath} = -\bar{z}\bar{x}$. En otras palabras, existe un entero racional e tal que $\bar{\imath} = \bar{e}$ o, equivalentemente, $i \equiv e \pmod{\pi}$. Esto hace que toda clase de k_π sea de la forma

$$\bar{a} + \bar{b}\bar{\imath} = \bar{a} + \bar{b}\bar{e} = \overline{a + be},$$

es decir, que todas las clases de restos de k_π es la clase de un entero, es decir, un elemento de $\mathbb{Z}_p$. Tenemos, pues, que $k_\pi = \mathbb{Z}_p$ y también se cumple en este caso que k_π tiene $\mathrm{N}(\pi) = p$ elementos.

En el ejemplo de $\pi = 2 + i$ tenemos trivialmente que $i \equiv -2 \pmod{\pi}$, por lo que, por ejemplo,

$$\overline{1 + 4i} = \bar{1} + \bar{4} \cdot (-\bar{2}) = \bar{3}.$$

En general, las clases de los enteros de las cuatro primeras filas de la tabla anterior coinciden con la clase de alguno de los de la última fila.

Observemos que si $\pi \equiv e \pmod{\pi}$, el número e no puede ser cualquiera, sino que esta congruencia implica que $-1 = \pi^2 \equiv e^2 \pmod{\pi}$, luego, según hemos probado, $-1 \equiv e^2 \pmod{p}$, luego $\bar{e}$ tiene que ser una de las raíces del polinomio $x^2 + 1$ en $\mathbb{Z}_p$ (en particular, si p se ramifica o se escinde, este polinomio tiene que tener raíces en $\mathbb{Z}_p$).

Además, π está determinado por p y e (salvo asociados), pues $\pi = (p, i - e)$.

En efecto, tenemos que $\pi \mid p$ y $\pi \mid i - e$, luego $\pi \mid (p, i - e)$, pero $p = \pi\bar{\pi}$, luego los únicos divisores de p múltiplos de π son π y p, pero $p \nmid i - e$ (pues $p \mid i - e$ implica que $p \mid e$ y $p \mid 1$), luego necesariamente $(p, i - e) = \pi$.

Recíprocamente, si $\bar{e}$ es una raíz de $x^2 + 1$ en $\mathbb{Z}_p$, entonces $\pi = (p, i - e)$ es un divisor primo de p. En efecto, tenemos que $x^2 + \bar{1} \equiv (x - \bar{e})(x - \bar{e}')$ en $\mathbb{Z}_p$, lo que equivale a que

$$(x - e)(x - e') - (x^2 + 1) = p\,q(x),$$

para cierto polinomio $q(x)$ en $\mathbb{Z}[x]$. Por lo tanto, $(i - e)(i - e') = p\,q(i)$, luego $p \mid (i - e)(i - e')$, pero p no puede dividir a ninguno de los dos factores, luego p no es primo, luego se descompone en dos factores primos (asociados o no) que no pueden dividir ambos al mismo factor (o éste sería múltiplo de p), luego cada uno de ellos divide a un factor distinto, luego existe un primo $\pi \mid p$ tal que $\pi \mid (i - e)$, luego $\pi \mid (p, i - e)$ y, como el máximo común divisor no puede ser p, de hecho $\pi = (p, i - e)$.

El teorema siguiente recoge lo que hemos obtenido:

Teorema 4.6 *Sea p un primo racional. Entonces:*

1. *Si $x^2+\bar{1} = (x-\bar{c})^2$ en $\mathbb{Z}_p[x]$, entonces $p = \epsilon\pi^2$, donde $\pi = (p, i-c)$ es un primo de Gauss que cumple $i \equiv c \pmod{\pi}$.*
2. *Si $x^2+\bar{1} = (x-\bar{c}_1)(x-\bar{c}_2)$ en $\mathbb{Z}_p[x]$, donde $c_1 \not\equiv c_2 \pmod{p}$, entonces $p = \pi_1\pi_2$, donde $\pi_i = (p, i-c_i)$ son primos de Gauss no asociados tales que $i \equiv c_j \pmod{\pi_j}$.*
3. *Si $x^2+\bar{1}$ es irreducible en $\mathbb{Z}_p[x]$, entonces p se conserva primo e i no es congruente módulo p con ningún entero racional.*

DEMOSTRACIÓN: Si $x^2+\bar{1}$ es irreducible en $\mathbb{Z}_p[x]$, entonces $\bar{i}$ no está en $\mathbb{Z}_p$, pues es una raíz de dicho polinomio, y hemos visto que eso implica que p se conserva primo en $\mathbb{Z}[i]$. Esto prueba 3).

Si $x^2+\bar{1} = (x-\bar{c})^2$ en $\mathbb{Z}_p$, hemos visto que $\pi = (p, i-c)$ es un divisor primo de p, así como que todo divisor primo de p tiene que ser de esta forma, luego π es el único divisor primo de p (salvo asociación). Esto significa que p se ramifica en $\mathbb{Z}[i]$.

Si $x^2+\bar{1} = (x-\bar{c}_1)(x-\bar{c}_2)$, hemos visto que $\pi_j = (p, i-c_j)$ son divisores primos de p, y no son asociados, pues si lo fueran sería $c_1 \equiv i \equiv c_2 \pmod{\pi_1}$, luego $c_1 \equiv c_2 \pmod{p}$. Por lo tanto, p se escinde en $\mathbb{Z}[i]$. ∎

Así pues, para determinar si un primo racional p se ramifica, se escinde o se conserva en $\mathbb{Z}[i]$, sólo tenemos que estudiar el comportamiento del polinomio x^2+1 en $\mathbb{Z}_p[x]$. Si tiene una única raíz es que p se ramifica, si tiene dos es que se escinde y si no tiene ninguna es que se conserva.

Ahora bien, para que el polinomio tenga una única raíz en $\mathbb{Z}_p$, con p impar, la condición necesaria y suficiente es que su discriminante $\Delta = -4$ sea 0 en $\mathbb{Z}_p$, lo cual no sucede nunca, luego:

El único primo racional que se ramifica en $\mathbb{Z}[i]$ es $2 = -i(1+i)^2$.

En el caso de un primo impar p, tenemos que p se escinde o se conserva según si $\Delta = -4$ es un cuadrado o no en $\mathbb{Z}_p$, lo cual equivale obviamente a que -1 sea o no un cuadrado en $\mathbb{Z}_p$. En este punto basta recordar el criterio de Euler 3.38 para concluir:

Teorema 4.7 *Si p es un primo racional entonces:*

- *p se ramifica en $\mathbb{Z}[i]$ si y sólo si $p = 2$.*
- *p se escinde en $\mathbb{Z}[i]$ si y sólo si $p \equiv 1 \pmod{4}$.*
- *p se conserva en $\mathbb{Z}[i]$ si y sólo si $p \equiv -1 \pmod{4}$.*

Y de aquí obtenemos a su vez la caracterización de los primos que son suma de dos cuadrados que obtuvimos en la sección anterior con un argumento distinto.

Nota Observemos que $U_4 = \{\pm 1\}$ y que, de los dos elementos de U_4, sólo 1 es un cuadrado, luego, para un primo impar, decir que $p \equiv 1 \pmod 4$ equivale a decir que p es un resto cuadrático módulo 4.

Por otra parte, aunque es más simple expresar la condición sobre los primos que se escinden en términos del -1, es interesante recordar que procede de exigir que el discriminante $\Delta = -4$ sea un resto cuadrático módulo p. Teniendo en cuenta estas consideraciones, podemos reformular en estos términos lo que hemos probado:

> *Si p es un primo impar, -4 es un resto cuadrático módulo p si y sólo si p es un resto cuadrático módulo -4.*

Obviamente, un resto módulo -4 es lo mismo que módulo 4, pero enunciado así podemos contemplar el caso más simple de lo que se conoce como "reciprocidad cuadrática". Volveremos sobre esto más adelante. ∎

En la prueba del teorema 4.6 hemos demostrado una parte del teorema siguiente:

Teorema 4.8 *Si α es un entero de Gauss no nulo, el anillo de clases de restos módulo α tiene $\mathrm{N}(\alpha)$ elementos.*

Demostración: Hemos probado el teorema para el caso en que α es primo, y es trivial si α es una unidad (pues entonces $\mathrm{N}(\alpha) = 1$ y, en efecto, todos los enteros de Gauss son congruentes módulo α, luego sólo hay una clase de congruencia).

Supongamos ahora que $\alpha = \pi^e$ es potencia de primo y razonamos por inducción sobre e. Lo tenemos probado para $e = 0, 1$. Supuesto cierto para e, sea $m = \mathrm{N}(\pi^e)$ y sea $n = \mathrm{N}(\pi)$, de modo que $\mathrm{N}(\pi^{e+1}) = mn$. Por hipótesis de inducción existen enteros de Gauss $\beta_1, \ldots, \beta_m$ tales que todo entero de Gauss es congruente módulo π^e con un único β_i, y también existen $\gamma_1, \ldots, \gamma_n$ tales que todo entero de Gauss es congruente módulo π con un único γ_j.

Basta probar que todo entero de Gauss η es congruente módulo π^{e+1} con un único entero de la forma $\beta_i + \pi^e\gamma_j$. En efecto, existe un i tal que

$$\eta = \beta_i + \pi^e\gamma$$

y, a su vez, existe un j tal que $\gamma = \gamma_j + \pi\delta$. Así

$$\eta = \beta_i + \pi^e(\gamma_j + \pi\delta) = \beta_i + \pi^e\gamma_j + \pi^{e+1}\delta \equiv \beta_i + \pi^e\gamma_j \pmod{\pi^{e+1}}.$$

Estos enteros no son congruentes entre sí, pues si

$$\beta_i + \pi^e\gamma_j \equiv \beta_{i'} + \pi^e\gamma_{j'} \pmod{\pi^{e+1}},$$

entonces

$$\beta_i + \pi^e\gamma_j = \beta_{i'} + \pi^e\gamma_{j'} + \delta\pi^{e+1},$$

luego $\pi^e \mid \beta_i - \beta_{i'}$, luego $i = i'$, por la unicidad de los β_i, luego

$$\pi^e\gamma_j = \pi^e\gamma_{j'} + \delta\pi^{e+1},$$

luego $\gamma_j = \gamma_{j'} + \delta\pi$, luego $j = j'$ por la unicidad de los γ_j.

Puesto que todo entero de Gauss se expresa en la forma $\alpha = \epsilon\pi_1^{e_1}\cdots\pi_r^{e_r}$, donde los π_j son primos no asociados dos a dos y ϵ es una unidad, y las potencias $\pi_i^{e_i}$ son primas entre sí dos a dos, el caso general se sigue del teorema siguiente. ∎

Teorema 4.9 (Teorema chino del resto) *Si $\alpha_1,\dots,\alpha_r$ son elementos de un dominio euclídeo A primos entre sí dos a dos y $\alpha = \alpha_1\cdots\alpha_r$, entonces la aplicación*

$$A_\alpha \longrightarrow A_{\alpha_1}\times\cdots\times A_{\alpha_r}$$

dada por $[\beta]_\alpha \mapsto ([\beta]_{\alpha_1},\dots,[\beta]_{\alpha_r})$ hace corresponder biunívocmente las clases de restos módulo α y las r-tuplas de clases de restos módulo los α_i.

Podríamos haber enunciado este teorema exactamente en los mismos términos que 3.7, en términos de congruencias, pero hemos optado por esta formulación equivalente.

Demostración: Es claro que

$$\beta \equiv \beta' \pmod{\alpha} \quad \text{si y sólo si} \quad \beta \equiv \beta' \pmod{\alpha_i} \quad \text{para todo } i,$$

lo que prueba que la aplicación está bien definida (no depende de la elección del entero β en cada clase de congruencia), así como que clases distintas tienen imágenes distintas. Que toda r-tupla es imagen de alguna clase $[\beta]_\alpha$ puede probarse adaptando literalmente el procedimiento explicado en la página 130, que en realidad es válido en dominios euclídeos cualesquiera. ∎

Terminamos con una observación elemental:

En la práctica, si sabemos que $i \equiv e \pmod{\pi}$, es fácil determinar si π divide o no a un entero de Gauss, pues

$$\pi \mid a+bi \quad \text{si y sólo si} \quad a+bi \equiv 0 \pmod{\pi} \quad \text{si y sólo si} \quad a+be \equiv 0 \pmod{\pi}$$

$$\text{si y sólo si} a+be \equiv 0 \pmod{p} \quad \text{si y sólo si} \quad p \mid a+be.$$ ∎

Ejemplo Consideremos $\alpha = 8+i$. Como $\mathrm{N}(\alpha) = 65 = 5\cdot 13$, al descomponer α en primos tiene que aparecer un primo de norma 5, es decir, uno de los dos primos $\pi_1 = 2+i$ o $\pi_2 = 2-i$ que dividen a 5. Para saber cuál de los dos, podemos realizar las divisiones:

$$\frac{8+i}{2+i} = \frac{(8+i)(2-i)}{5} = \frac{17}{5} - \frac{6}{5}i, \qquad \frac{8+i}{2-i} = \frac{(8+i)(2+i)}{5} = 3+2i,$$

y así vemos que es $2-i$ el primo que divide a α. Pero una forma alternativa de comprobarlo es observar que $i \equiv -2 \pmod{\pi_1}$, mientras que $i \equiv 2 \pmod{\pi_2}$. Por lo tanto,

$$\alpha = 8+i \equiv 8-2 = 6 \equiv 1 \pmod{\pi_1}, \qquad \alpha = 8+i \equiv 8+2 \equiv 0 \pmod{\pi_2},$$

luego también concluimos así que es π_2. ∎

4.3 Sumas de dos cuadrados II

Vamos a resolver de nuevo el problema sobre los números naturales que pueden expresarse como sumas de dos cuadrados, pero usando ahora la herramienta que hemos desarrollado en la sección precedente. En primer lugar, ahora podemos reformular el problema en estos términos:

¿Qué números naturales son normas de enteros de Gauss?

El teorema 4.7 nos da la respuesta para los primos:

Un primo p es suma de dos cuadrados si y sólo si $p = 2$ o cumple $p \equiv 1 \pmod 4$.

Supongamos ahora que nos dan un número natural n y queremos saber si es o no suma de dos cuadrados. Sabemos que esto equivale a que exista un entero de Gauss α tal que $\mathrm{N}(\alpha) = n$. Veamos qué condiciones tienen que cumplirse para que podamos encontrar tal entero α.

Dado, en principio, un α arbitrario, admitirá una descomposición en primos de Gauss, digamos

$$\epsilon\lambda^{e_0} p_1^{e_1} \cdots p_k^{e_k} \pi_{k+1}^{e_{k+1}} \cdots \pi_r^{e_r},$$

donde ϵ es una unidad y hemos distinguido el primo $\lambda = 1 + i$ de norma 2, los primos p_j de norma p_j^2 y los primos π_j de norma prima p_j. Al calcular la norma queda

$$n = \mathrm{N}(\alpha) = 2^{e_0} p_1^{2e_1} \cdots p_k^{2e_k} p_{k+1}^{e_{k+1}} \cdots p_r^{e_r},$$

donde $p_1, \ldots, p_k$ son primos que se conservan (es decir, primos $p_j \equiv -1 \pmod 4$) y $p_{k+1}, \ldots, p_r$ son primos que se ramifican.

Vemos que la única restricción para que un número natural n sea de esta forma, es decir, que sea la norma de un entero de Gauss, es que los primos $p \equiv -1 \pmod 4$ que dividan a n tienen que aparecer con exponente par. Equivalentemente, que todos los primos p que dividan a n con exponente impar tienen que ser $p = 2$ o bien $p \equiv 1 \pmod 4$. Así pues:

Un número natural n es suma de dos cuadrados si y sólo si los primos p que lo dividen con exponente impar son $p = 2$ o $p \equiv 1 \pmod 4$.

En particular, esto implica que un número es suma de dos cuadrados si y sólo si lo es su parte libre de cuadrados, y en general todo lo que habíamos obtenido en la sección 4.1 por medios más laboriosos y rudimentarios.

Más aún, con el análisis que hemos hecho, podemos responder fácilmente a una pregunta más fina:

¿De cuántas formas distintas puede expresarse un número natural como suma de dos cuadrados?

Por ejemplo:

$$\begin{aligned} 21\,125 = 5^3 \cdot 13^2 &= 10^2 + 145^2 = 26^2 + 43^2 = 31^2 + 142^2 \\ &= 65^2 + 130^2 = 79^2 + 122^2 = 95^2 + 110^2, \end{aligned}$$

y no hay más posibilidades, salvo reordenación de los sumandos.

En principio, contar formas de expresar un número n como suma de dos cuadrados equivale a contar cuántos enteros de Gauss hay con norma n, pero esto no es exacto. Por ejemplo, si $\alpha = 10 + 145i$, tanto sus asociados como sus conjugados dan "la misma expresión":

$$\begin{array}{cccc} 10+145i & -145+10i & -10-145i & 145-10i \\ 10-145i & -145-10i & -10+145i & 145+10i \end{array}$$

y claramente no hay más posibilidades que den "la misma expresión". Veamos en primer lugar que es fácil contar cuántas clases de enteros de Gauss asociados tienen una norma n dada. Para ello descomponemos n en factores primos:

$$n = 2^{e_0} p_1^{e_1} \cdots p_k^{e_k} p_{k+1}^{e_{k+1}} \cdots p_r^{e_r},$$

donde separamos $p_0 = 2$, los primos $p_j \equiv -1$ (mód 4) (para $1 \le j \le k$) y los primos $p_j \equiv 1$ (mód 4) (para $k+1 \le j \le r$). Según hemos visto, la condición necesaria y suficiente para que existan enteros de Gauss α de norma n es que los exponentes e_j (para $1 \le j \le k$) sean pares. Admitiendo que esto es así, α es necesariamente de la forma

$$\alpha = \epsilon \lambda^{e_0} p_1^{e_1/2} \cdots p_k^{e_k/2}\ \pi_{k+1}^{s_{k+1}}\ \bar{\pi}_{k+1}^{e_{k+1}-s_{k+1}} \cdots \pi_r^{s_r}\ \bar{\pi}_r^{e_r-s_r},$$

donde ϵ es una unidad y hemos elegido arbitrariamente $\lambda = 1+i$ como primo de norma 2, cada p_j como primo de norma p_j^2 (para $1 \le j \le k$) y uno cualquiera de los divisores primos π_j de p_j (para $1 \le j \le k$).

Si fijamos, por ejemplo, $\epsilon = 1$, los α construidos de este modo serán no asociados dos a dos (pues no dividirán exactamente a los mismos enteros de Gauss) y nuestro único margen de libertad consiste en obtener las potencias $p_j^{e_j}$ (para $1 \le j \le k$) empleando combinaciones distintas de π_j y $\bar{\pi}_j$, lo que nos permite elegir cada s_j entre 0 y e_j, luego nos da $e_j + 1$ posibilidades por cada j. En total, el número de enteros de Gauss de norma

$$n = 2^{e_0} p_1^{e_1} \cdots p_k^{e_k} p_{k+1}^{e_{k+1}} \cdots p_r^{e_r}$$

sin contar cuatro veces cada clase de asociados (y supuesto que e_j es par para $1 \le j \le k$) es

$$N_0 = (e_{k+1} + 1) \cdots (e_r + 1),$$

pero este número hay que dividirlo entre 2 para no contar las representaciones asociadas a cada α y su conjugado $\bar{\alpha}$ como dos representaciones distintas, salvo que α cumpla $\bar{\alpha} = \alpha$, en cuyo caso la representación asociada no la estamos contando dos veces y no hay nada que corregir.

Veamos cuándo se da este caso peculiar. Ante todo, la condición $\bar{\alpha} = \alpha$ equivale a que α sea un entero racional, digamos $\alpha = x$ y sus cuatro asociados $\pm\alpha$, $\pm i\alpha$, dan lugar a la representación $n = x^2 + 0^2$.

Así pues, para que se dé este caso, es necesario (y claramente suficiente) que n sea un cuadrado perfecto. Pero si $n = x^2$, entonces la única representación posible asociada a un α entero racional es precisamente la correspondiente a $\alpha = x$ (o a un asociado a x), por lo que el caso excepcional sólo puede darse para un único α.

Concluimos entonces que el número de formas distintas de representar n como suma de dos cuadrados es $N_0/2$ si n no es un cuadrado perfecto, o bien $(N_0 + 1)/2$ si es que lo es. (Tenemos $N_0 - 1$ representaciones contadas por duplicado y otra que no, luego en total hay $1 + (N_0 - 1)/2$ representaciones.)

Enunciamos a continuación el resultado que hemos obtenido:

> *Un número natural n es expresable como suma de dos cuadrados si y sólo si sus divisores primos $p \equiv -1 \pmod 4$ lo dividen con exponente par y, en tal caso, si $e_1, \dots, e_r$ son los exponentes de sus divisores primos $p \equiv 1 \pmod 4$ y $N_0 = (e_1+1)\cdots(e_r+1)$, el número de formas distintas en que puede ser expresado como suma de dos cuadrados es $N_0/2$ si n no es un cuadrado perfecto y $(N_0+1)/2$ en caso contrario.*

Por ejemplo, en el caso de $21\,125 = 5^3 \cdot 13^2$, tenemos $N_0 = (3+1)(2+1) = 12$ luego hay 6 formas de expresarlo como suma de dos cuadrados. En el caso de $4\,225 = 5^2 \cdot 13^2$ es $N_0 = (2+1)(2+1) = 9$ y hay 5 formas de expresarlo como suma de dos cuadrados.

Ejercicio: Construir las 5 representaciones de 4 225 como suma de dos cuadrados como normas de los enteros de Gauss adecuados.

4.4 Otras aplicaciones de los enteros de Gauss

Ternas pitagóricas Vamos a ver que, usando enteros de Gauss, la parametrización de las ternas pitagóricas primitivas se simplifica ligeramente.

Partimos de una terna pitagórica primitiva $x^2 + y^2 = z^2$. En particular cumple que $(x, y) = 1$. Consideremos el entero de Gauss $\alpha = x + yi$. El hecho de que $(x, y) = 1$ implica que α no es divisible entre enteros racionales no unitarios, pues si $k \mid \alpha$, entonces $x + yi = k(u + vi)$, luego $k \mid x$, $k \mid y$.

En particular, α no puede ser divisible por primos de norma p^2 y, si un primo π de norma p divide a α, no puede suceder que $\bar{\pi}$ también lo divida, pues entonces $p = \pi\bar{\pi}$ dividiría a α. Esto significa que

$$\alpha = \epsilon\pi_1^{e_1} \cdots \pi_r^{e_r},$$

donde ϵ es una unidad, los primos π_i tienen todos norma prima y además no hay dos con la misma norma. Así pues, $z^2 = x^2 + y^2 = \mathrm{N}(\alpha) = p_1^{e_1} \cdots p_r^{e_r}$, donde los primos p_i son distintos dos a dos. La factorización única de $\mathbb{Z}$ implica que todos los exponentes son pares. Por lo tanto, llamando $p + qi = \pi_1^{e_1/2} \cdots \pi_r^{e_r/2}$, tenemos que

$$\alpha = x + yi = \epsilon(p + qi)^2 = \epsilon(p^2 - q^2 + 2pqi).$$

Si $\epsilon = -1$ o $\epsilon = -i$, podemos introducir el $-1 = i^2$ en el cuadrado, con lo que podemos suponer que $\epsilon = 1, i$, pero si $\epsilon = i$, intercambiando x e y podemos suponer que $\epsilon = 1$. En definitiva, $x = p^2 - q^2$, $y = 2pq$ y

$$z^2 = (p + iq)^2(p - iq)^2 = (p^2 + q^2)^2,$$

luego si suponemos que x, y, $z > 0$, tiene que ser

$$(x, y, z) = (p^2 - q^2, 2pq, p^2 + q^2)$$

con $q < p$ y $(p, q) = 1$ (para que se cumpla $(x, y) = 1$) y de paridad opuesta (para que x sea impar). ∎

Ecuaciones de Mordell Vamos a usar los enteros de Gauss para resolver algunas ecuaciones de Mordell. La primera que vamos a estudiar la resolvió Fermat (sin usar enteros de Gauss, naturalmente):

Las únicas soluciones enteras de la ecuación $y^2 + 4 = x^3$ *son*

$$(x, y) = (5, \pm 11),\ (2, \pm 2).$$

Demostración: Supongamos primero que y es impar. Tenemos que

$$(2 + iy)(2 - iy) = x^3.$$

Un divisor común α de $2 + iy$, $2 - iy$ lo es también de su suma, 4, luego $\mathrm{N}(\alpha) \mid 16$ y $\mathrm{N}(\alpha) \mid x^3$, impar, lo que implica que $\alpha = 1$ y los dos factores $2 + iy$, $2 - iy$ son primos entre sí. De aquí podemos deducir que los dos factores son cubos, pero es importante destacar que aquí es fundamental que todas las unidades de $\mathbb{Z}[i]$ son cubos.[2] Así pues, existen enteros racionales a y b tales que $2 + iy = (a + ib)^3$. Conjugando, $2 - iy = (a - ib)^3$, y sumando las dos ecuaciones y simplificando se llega a que $4 = 2a(a^2 - 3b^2)$, luego tenemos que $a(a^2 - 3b^2) = 2$.

Claramente entonces $a = \pm 1, \pm 2$, pero las únicas posibilidades que permiten la existencia de b son $a = -1$, $b = \pm 1$ y $a = 2$, $b = \pm 1$. Entonces

$$x^3 = \big((a + ib)(a - ib)\big)^3 = (a^2 + b^2)^3$$

y los valores de x que se obtienen son $x = 2, 5$. De $y^2 + 4 = 8, 125$, obtenemos que $y = \pm 2, \pm 11$.

[2]En principio, lo que sabemos es que todos los divisores primos de cada factor tienen exponente múltiplo de 3, lo cual significa que los factores son de la forma $\epsilon\pi_1^{3e_1} \cdots \pi_r^{3e_r}$, pero para concluir que son cubos es necesario observar además que $\epsilon = 1, -1, i - 1$ puede ponerse de la forma 1^3, $(-1)^3$, $(-i)^3$ o i^3, respectivamente.

Ahora queda el caso en que y es par. Digamos $y = 2Y$. Entonces x ha de ser par también, $x = 2X$ y se cumple $Y^2 + 1 = 2X^3$. De aquí se sigue que Y es impar. Factorizamos

$$(1 + iY)(1 - iY) = 2X^3. \tag{4.1}$$

El máximo común divisor de $1 + iY$ y $1 - iY$ divide a la suma $2 = -i(1 + i)^2$. Como $1 + i$ es primo (tiene norma 2), dicho mcd ha de ser 1, $1 + i$, o bien 2.

Claramente 2 no divide a $1 + iY$, sin embargo, $1 + i$ sí divide tanto a $1 + iY$ como a $1 - iY$ (se sigue de que Y es impar).

Al dividir los dos miembros de (4.1) entre $(1 + i)^2$, en el primer miembro queda un producto de dos factores primos entre sí, y en el segundo miembro queda $-iX^3 = (iX)^3$. Por lo tanto, cada factor del primer miembro ha de ser un cubo. En particular $(1+iY)/(1+i)$ es un cubo, es decir, $1+iY = (1+i)(a+ib)^3$. Conjugando y sumando como antes se llega a $1 = (a + b)(a^2 - 4ab + b^2)$ y de aquí salen las soluciones $a = \pm 1$, $b = 0$ o bien $a = 0$, $b = \pm 1$, que implican $y = \pm 2$, y entonces $x = 2$. ∎

En el capítulo anterior resolvimos con gran esfuerzo la ecuación $y^2 = x^3 + 1$. Veamos ahora que el problema de qué cuadrados preceden a cubos es bastante más fácil de resolver:

Ejemplo *La única solución entera de la ecuación* $y^2 = x^3 - 1$ *es* $(x, y) = (1, 0)$.

Demostración: Si x es par resulta que $y^2 \equiv -1$ (mód 8), lo cual es imposible, luego x tiene que ser impar y, por consiguiente, y es par. Podemos factorizar:

$$x^3 = (y + i)(y - i).$$

Los dos factores son primos entre sí, pues un divisor primo común tiene que dividir a su diferencia, $2i$, luego tiene que ser $\lambda = 1+i$, pero entonces λ dividiría a x, que es impar. Por consiguiente, y teniendo en cuenta que las unidades de $\mathbb{Z}[i]$ son cubos, ambos factores son cubos. Pongamos que

$$y + i = (m + ni)^3.$$

Desarrollando el cubo e igualando los coeficientes, obtenemos que

$$y = m^3 - 3mn^2 = m(m^2 - 3n^2), \quad 1 = 3m^2n - n^3 = n(3m^2 - n^2).$$

La ecuación de la derecha implica que $n = \pm 1$. Si $n = 1$ se reduce a $3m^2 - 1 = 1$, que no tiene solución entera. Por lo tanto tiene que ser $n = -1$, y así la ecuación es $3m^2 - 1 = -1$ nos da que $m = 0$, luego la ecuación de la izquierda nos da que $y = 0$, luego la ecuación de partida nos da que $x = 1$. ∎

El argumento se puede generalizar:

La única solución entera de $y^2 + 1 = x^p$ *con* $p \geq 2$ *es* $(1, 0)$.

Demostración: Si existiera una solución con $y \neq 0$ y $p = p'm$, con p' primo, entonces (x^m, y) sería una solución no trivial de la ecuación $y^2 + 1 = x^{p'}$, luego podemos suponer que p es primo. Es obvio que dos cuadrados no pueden

ser consecutivos, así que podemos suponer que $p \geq 3$. Si x es par concluimos como antes que $y^2 \equiv -1 \pmod 8$, lo cual es imposible, luego x es impar y esto hace que y sea par. Factorizamos

$$x^p = (y+i)(y-i).$$

Los factores son primos entre sí por el mismo argumento empleado en el ejemplo anterior. Las unidades de $\mathbb{Z}[i]$ son potencias p-ésimas, pues $i^p = \pm i$ según si $p \equiv \pm 1 \pmod 4$, con lo que $(-i)^p = \mp i$. Por lo tanto, ambos factores son potencias p-ésimas. Pongamos que

$$y + i = (m + ni)^p = \sum_{j=0}^{p} \binom{p}{j} m^j (bi)^{p-j}.$$

Igualando las partes imaginarias queda

$$1 = \sum_{j=0}^{(p-1)/2} \binom{p}{2j} m^{2j} n^{p-2j} i^{p-2j-1} = n \sum_{j=0}^{(p-1)/2} \binom{p}{2j} m^{2j} n^{p-2j-1} (-1)^{(p-1)/2-j}.$$

De aquí se sigue que $n = \pm 1$. Multiplicando por $(-1)^{(p-1)/2} n$ queda

$$\sum_{j=0}^{(p-1)/2} \binom{p}{2j} (-m^2)^j = (-1)^{(p-1)/2} n.$$

Por otra parte, $y - i = (m - ni)^p$, luego

$$x^p = (m+ni)^p (m-ni)^p = (m+i)^p (m-i)^p = (m^2+1)^p.$$

En particular m es par y no nulo (pues si fuera $m = 0$ entonces $x = 1$ e $y = 0$).

Por lo tanto

$$\sum_{j=0}^{(p-1)/2} \binom{p}{2j} (-m^2)^j \equiv 1 \pmod 4,$$

luego necesariamente

$$\sum_{j=0}^{(p-1)/2} \binom{p}{2j} (-m^2)^j = 1.$$

Si $p = 3$, esta ecuación implica que $m = 0$, luego $p \geq 5$, luego $(p-1)/2 \geq 2$. Por lo tanto,

$$\sum_{j=2}^{(p-1)/2} \binom{p}{2j} (-m^2)^j = m^2 \binom{p}{2}.$$

Terminaremos la prueba mostrando que el exponente de 2 en el miembro izquierdo de la ecuación anterior es estrictamente mayor que en el derecho. Para ello nos apoyamos en la igualdad siguiente:

$$\binom{p}{2j} = \frac{1}{j(2j-1)} \binom{p-2}{2j-2} \binom{p}{2},$$

válida para $j = 2, \ldots, (p-1)/2$.

La prueba es inmediata:

$$\frac{1}{j(2j-1)}\binom{p-2}{2j-2}\binom{p}{2} = \frac{1}{j(2j-1)}\frac{(p-2)!}{(2j-2)!(p-2j)!}\frac{p(p-1)}{2}$$
$$= \frac{p!}{(2j)!(p-2j)!} = \binom{p}{2j}.$$

Así:

$$v_2\left(m^{2j}\binom{p}{2j}\right) \geq 2jv_2(m) - v_2(j) + v_2\left(\binom{p}{2}\right)$$
$$\geq (2j - v_2(j))v_2(m) + v_2\left(\binom{p}{2}\right) > 2v_2(m) + v_2\left(\binom{p}{2}\right).$$

Hemos usado que $2j - v_2(j) > 2$, pues esto equivale a que $v_2(j) < 2(j-1)$, y es cierto, pues claramente $v_2(j) < j$, luego $v_2(j) \leq j - 1 < 2(j-1)$. En conclusión, 2 divide a cada sumando del miembro izquierdo con exponente mayor que al miembro derecho, luego lo mismo vale para toda la suma y tenemos una contradicción. ∎

El ejemplo siguiente no es realmente una aplicación de los enteros de Gauss, sino de teorema 3.38:

Ejemplo *La ecuación $y^2 = x^3 - 5$ no tiene soluciones enteras.*

DEMOSTRACIÓN: Consideremos los dos miembros de la ecuación módulo 4:

y	y^2	x	x^3-1
0	0	0	3
1	1	1	0
2	0	2	3
3	1	3	2

Vemos que para que se dé la igualdad es necesario que $x \equiv 1$ (mód 4) y que $y \equiv 0, 2$ (mód 4), lo cual significa que y tiene que ser par. Escribimos la ecuación en la forma

$$y^2 + 4 = x^3 - 1 = (x-1)(x^2 + x + 1).$$

Observemos que $x^2 + x + 1 = (x - 1/2)^2 + 3/2 > 0$, y además

$$x^2 + x + 1 \equiv 1 + 1 + 1 \equiv -1 \text{ (mód 4)}.$$

Si descomponemos $x^2 + x + 1$ en factores primos (positivos), alguno de ellos p tiene que cumplir $p \equiv -1$ (mód 4), ya que si todos fueran congruentes con 1 módulo 4 lo mismo le sucedería al producto de todos ellos, es decir, a $x^2 + x + 1$.

Entonces $p \mid y^2 + 4$, luego $y^2 \equiv -4$ (mód p). Expresando $y = 2k$ y simplificando el 4 queda $-1 \equiv k^2$ (mód p), es decir, que -1 es un resto cuadrático módulo p, en contra de lo que afirma el teorema 3.38. Por lo tanto, la ecuación no tiene soluciones enteras. ∎

Ejercicio: Probar que la ecuación $y^2 = x^3 + 11$ no tiene soluciones enteras. AYUDA: Razonar como en el ejemplo anterior, usando ahora la factorización

$$y^2 + 16 = x^3 + 27 = (x+3)(x^2 - 3x + 9).$$

Capítulo V

El símbolo de Legendre

En el capítulo anterior hemos visto que el teorema de Euler 3.38 sobre el carácter cuadrático de -1 módulo un primo arbitrario es la "razón última" que explica qué números pueden expresarse como suma de dos cuadrados, así como que puede usarse para determinar si determinadas ecuaciones tienen o no soluciones enteras. En este capítulo vamos a plantear problemas similares relacionados con el carácter cuadrático de -2.

El lector que haya asimilado las características del problema sobre los números expresables como suma de dos cuadrados debería ser capaz de conjeturar fácilmente la situación correspondiente a los números que pueden expresarse en la forma $n = x^2 + 2y^2$. La tabla siguiente recoge los primeros de ellos:

0	1	**2**	**3**	4	6	8	9	**11**	12	16	**17**	18
19	22	24	25	27	32	33	34	36	38	**41**	**43**	44
48	49	50	51	54	57	**59**	64	66	**67**	68	72	**73**
75	76	81	82	**83**	86	88	89	96	**97**	98	99	100

5.1 El anillo $\mathbb{Z}[\sqrt{-2}\,]$

Para estudiar el problema que acabamos de plantear necesitamos "diseñar" un anillo a medida, del mismo modo que el anillo $\mathbb{Z}[i]$ de los enteros de Gauss le venía "como anillo al dedo" al problema de las sumas de dos cuadrados. Queremos un anillo que nos permita factorizar:

$$x^2 + 2y^2 = (x + \sqrt{-2})(x - \sqrt{-2}).$$

La diferencia es que en el capítulo anterior queríamos una raíz cuadrada de -1 y ahora queremos una raíz cuadrada de -2. Para obtenerla sólo tenemos que imitar la construcción del cuerpo $\mathbb{Q}(i)$ cambiando el polinomio x^2+1 por x^2+2.

En efecto, el polinomio $x^2 + 2$ es irreducible en $\mathbb{Q}[x]$, porque tiene grado 2 y no tiene raíces en $\mathbb{Q}$. Por lo tanto, el teorema 4.1 nos asegura que el anillo de clases de restos de $\mathbb{Q}[x]$ módulo $x^2 + 2$ es un cuerpo, al que de momento llamaremos k.

En principio, sus elementos son clases de restos $\overline{q(x)}$, donde $q(x)$ es un polinomio arbitrario, de $\mathbb{Q}(x)$, pero realizando la división euclídea

$$q(x) = (x^2+1)c(x) + a + bx,$$

concluimos que los elementos de k son de la forma $\bar{a} + \bar{b}\bar{x}$, donde a y b son números racionales. Además, la expresión es única. El lector debería completar los detalles de este argumento —y de los siguientes— sin más que comparar en caso de duda con los dados en el capítulo anterior en la construcción del cuerpo $\mathbb{Q}(i)$.

La unicidad de la expresión hace que podamos identificar cada número racional a con la clase $\bar{a}$ de k, de modo que podemos considerar que el cuerpo k contiene a $\mathbb{Q}$, y sus elementos se expresan (de forma única) en la forma $a + b\bar{x}$, donde a y b son números racionales. Finalmente observamos que $\overline{x^2+2} = \bar{0}$, lo que equivale a que $\bar{x}^2 = -2$. Ello nos lleva a definir $\sqrt{-2} = \bar{x}$.

Definición 5.1 Llamaremos $\mathbb{Q}\big(\sqrt{-2}\,\big)$ al cuerpo de clases de restos del anillo de polinomios $\mathbb{Q}[x]$ módulo el polinomio x^2+2. Llamando $\sqrt{-2} = \bar{x}$ e identificando cada número racional a con la clase $\bar{a}$, tenemos que cada elemento de $\mathbb{Q}\big(\sqrt{-2}\,\big)$ se expresa de forma única como $a + b\sqrt{-2}$ con a y b números racionales, y además se cumple que $(\sqrt{-2})^2 = -2$.

Estos hechos determinan completamente el álgebra de $\mathbb{Q}\big(\sqrt{-2}\,\big)$. Sus elementos se suman y se multiplican así:

$$\begin{aligned}(a+b\sqrt{-2}) + (c+d\sqrt{-2}) &= (a+c) + (b+d)\sqrt{-2},\\ (a+b\sqrt{-2})(c+d\sqrt{-2}) &= ac + bd(\sqrt{-2})^2 + ad\sqrt{-2} + bc\sqrt{-2}\\ &= ac - 2bd + (ad+bc)\sqrt{-2}\end{aligned}$$

Por último definimos el *conjugado* de un número $a + b\sqrt{-2}$ como

$$\overline{a + b\sqrt{-2}} = a - b\sqrt{-2},$$

y se cumple que

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \qquad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2.$$

Nuevamente, estas relaciones pueden probarse calculando explícitamente los dos miembros de cada una, y los cálculos serían análogos a los correspondientes al cuerpo $\mathbb{Q}(i)$, pero el razonamiento abstracto alternativo que vimos en el capítulo anterior vale sin cambio alguno en este caso sin cálculo alguno: definimos la aplicación

$$c : \mathbb{Q}[x] \longrightarrow \mathbb{Q}\big(\sqrt{-2}\,\big)$$

dada por $c(p(x)) = p(-\sqrt{-2})$, de modo que se cumple trivialmente que

$$c(p(x) + q(x)) = c(p(x)) + c(q(x)),$$

$$c(p(x)q(x)) = c(p(x))x(q(x)).$$

Resumen 5.1: El cuerpo $\mathbb{Q}\big(\sqrt{-2}\,\big)$

- $\mathbb{Q}\big(\sqrt{-2}\,\big)$ es un cuerpo cuyos elementos se expresan de forma única como $a+b\sqrt{-2}$, donde a y b son números racionales y $(\sqrt{-2})^2=-2$.
- La suma y el producto en $\mathbb{Q}\big(\sqrt{-2}\,\big)$ vienen dadas por

$$\begin{aligned}(a+b\sqrt{-2})+(c+d\sqrt{-2}) &= (a+c)+(b+d)\sqrt{-2},\\ (a+b\sqrt{-2})(c+d\sqrt{-2}) &= ac-2bd+(ad+bc)\sqrt{-2}\end{aligned}$$

- El *conjugado* de un elemento de $\mathbb{Q}\big(\sqrt{-2}\,\big)$ es $\overline{a+b\sqrt{-2}}=a-b\sqrt{-2}$.
- La conjugación cumple las propiedades

$$\overline{z_1+z_2}=\bar{z}_1+\bar{z}_2,\qquad \overline{z_1z_2}=\bar{z}_1\bar{z}_2.$$

- La *norma* $\mathrm{N}:\mathbb{Q}\big(\sqrt{-2}\,\big)\longrightarrow\mathbb{Q}\big(\sqrt{-2}\,\big)$ se define como

$$\mathrm{N}(z)=z\bar{z},\qquad \text{o equivalentemente.}\qquad \mathrm{N}(a+bi)=a^2+2b^2.$$

- Se cumple que $\mathrm{N}(z_1z_1)=\mathrm{N}(z_1)\,\mathrm{N}(z_2)$.
- Si $z\neq 0$ es un elemento de $\mathbb{Q}\big(\sqrt{-2}\,\big)$, entonces

$$z^{-1}=\frac{\bar{z}}{\mathrm{N}(z)}.$$

Además, $c(x^2+2)=(-\sqrt{-2})^2+2=0$, luego si $p(x)\equiv q(x)\pmod{x^2+2}$, esto significa que $p(x)-q(x)=(x^2+2)h(x)$ y, evaluando en $-\sqrt{-2}$, queda que $c(p(x))-c(q(x))=0$, luego $c(p(x))=c(q(x))$. Por consiguiente, c induce una aplicación

$$\bar{c}:\mathbb{Q}\big(\sqrt{-2}\,\big)\longrightarrow\mathbb{Q}\big(\sqrt{-2}\,\big)$$

dada por $\bar{c}([p(x)])=c(p(x))$, que está bien definida porque hemos visto que no importa qué polinomio $p(x)$ tomamos como representante de la clase de restos. Es inmediato que $\bar{c}$ cumple también

$$\bar{c}(z_1+z_2)=\bar{c}(z_1)+\bar{c}(z_2),\qquad \bar{c}(z_1z_2)=\bar{c}(z_1)\bar{c}(z_2)$$

y claramente $\bar{c}$ no es sino la conjugación que hemos definido antes, pues

$$\bar{c}(a+b\sqrt{-2})=\bar{c}([a+bx])=a-b\sqrt{-2}.$$

A partir de aquí el lector puede olvidar que los elementos de $\mathbb{Q}\big(\sqrt{-2}\,\big)$ son clases de equivalencia de polinomios y recordar únicamente que cumple las propiedades básicas recogidas en el resumen 5.1, análogo al resumen 4.1. En él

viene la definición de una norma análoga a la que consideramos sobre los enteros de Gauss, junto con la fórmula que expresa el inverso de un elemento en términos de su conjugado y de su norma.

Ejercicio: Probar todas las propiedades contenidas en el resumen 5.1.

He aquí un ejemplo de operación en el cuerpo que acabamos de construir:

$$\frac{2+\sqrt{-2}}{3-4\sqrt{-2}} = \frac{(2+\sqrt{-2})(3+4\sqrt{-2})}{(3-4\sqrt{-2})(3+4\sqrt{-2})} = \frac{-2+11\sqrt{-2}}{41} = -\frac{2}{41} + \frac{11}{41}\sqrt{-2}.$$

Definición 5.2 Definimos el anillo $\mathbb{Z}\big[\sqrt{-2}\,\big]$ como el formado por los elementos de $\mathbb{Q}\big(\sqrt{-2}\,\big)$ de la forma $a+b\sqrt{-2}$, donde a y b son números enteros.

Es habitual referirse a los elementos de $\mathbb{Z}\big[\sqrt{-2}\,\big]$ como *enteros cuadráticos,* y por oposición llamar entonces *enteros racionales* a los enteros ordinarios.

Ejercicio: Probar que $\mathbb{Q}(\sqrt{-2})$ es el cuerpo de cocientes de $\mathbb{Z}[\sqrt{-2}]$.

Teorema 5.3 *El anillo* $\mathbb{Z}\big[\sqrt{-2}\,\big]$ *es un dominio euclídeo con la norma euclídea* $\mathrm{N} : \mathbb{Z}\big[\sqrt{-2}\,\big] \longrightarrow \mathbb{N}$.

Demostración: Teniendo en cuenta que las normas de los enteros cuadráticos son números naturales, es evidente que si α y β son enteros cuadráticos no nulos, entonces $\mathrm{N}(\alpha) \le \mathrm{N}(\alpha)\,\mathrm{N}(\beta) = \mathrm{N}(\alpha\beta)$.

Tomemos ahora dos enteros cuadráticos Δ y δ con $\delta \neq 0$. Consideramos el cociente $\Delta/\delta = r + s\sqrt{-2}$, donde r y s son números racionales. Llamamos x e y a los enteros racionales más cercanos a r y s, respectivamente, de modo que $|r-x|$, $|s-y| \le 1/2$, con lo que, llamando $\gamma = x + y\sqrt{-2}$, $\epsilon = \Delta - \delta\gamma$, tenemos que

$$\mathrm{N}(\epsilon/\delta) = \mathrm{N}(\Delta/\delta - \gamma) = (r-x)^2 + 2(s-y)^2 \le 1/4 + 2/4 = 3/4 < 1,$$

luego $\mathrm{N}(\epsilon) < \mathrm{N}(\delta)$. ∎

Con esto tenemos justificado que $\mathbb{Z}\big[\sqrt{-2}\,\big]$ tiene una aritmética con las mismas propiedades básicas que la de $\mathbb{Z}$ o $\mathbb{Z}[i]$. En particular es un dominio de factorización única. Por ejemplo, he aquí una aplicación del algoritmo de Euclides que muestra que $(35+80\sqrt{-2}, 13-3\sqrt{-2}) = 1$:

c	r	u	v	N
	$35+80\sqrt{-2}$	1	0	14 025
$6\sqrt{-2}$	$13-3\sqrt{-2}$	0	1	187
$-3-3\sqrt{-2}$	$-1+2\sqrt{-2}$	1	$-6\sqrt{-2}$	9
$-\sqrt{-2}$	-2	$3+3\sqrt{-2}$	$37-18\sqrt{-2}$	4
	-1	$-5+3\sqrt{-2}$	$36+31\sqrt{-2}$	1

Además nos da la relación de Bezout:

$$(5-3\sqrt{-2})(35+80\sqrt{-2}) - (36+31\sqrt{-2})(13-3\sqrt{-2}) = 1.$$

Unidades El mismo razonamiento empleado con los enteros de Gauss prueba que las unidades de $\mathbb{Z}\big[\sqrt{-2}\,\big]$ son los enteros que cumplen $\mathrm{N}(\alpha) = 1$. Pero en este caso, si $\alpha = x + y\sqrt{-2}$, la ecuación $\mathrm{N}(\alpha) = x^2 + 2y^2 = 1$ sólo tiene las soluciones $(x, y) = (\pm 1, 0)$. Por lo tanto:

Las unidades de $\mathbb{Z}\big[\sqrt{-2}\,\big]$ son ± 1.

Ésta es una primera muestra de que, aunque $\mathbb{Z}\big[\sqrt{-2}\,\big]$ cumpla las mismas propiedades aritméticas generales que $\mathbb{Z}$ o que $\mathbb{Z}[i]$, esto no impide que la aritmética de cada anillo tenga su "personalidad propia". Así, $\mathbb{Z}\big[\sqrt{-2}\,\big]$ se parece más a $\mathbb{Z}$ que a $\mathbb{Z}[i]$ en cuanto a que los asociados se agrupan por parejas y no por cuádruplas. De este modo, los asociados de $3 - 2\sqrt{-2}$ son únicamente él mismo y $-3 + 2\sqrt{-2}$.

Primos Si π es primo en $\mathbb{Z}\big[\sqrt{-2}\,\big]$, entonces $\pi \mid \pi\bar{\pi} = \mathrm{N}(\pi)$, por lo que, descomponiendo $\mathrm{N}(\pi)$ en primos racionales, concluimos que π divide a uno de ellos. Así pues, todo primo cuadrático aparece en la descomposición en factores primos de un primo racional.

Además, si $\pi \mid p$, tenemos que $\mathrm{N}(\pi) \mid \mathrm{N}(p) = p^2$, lo que prueba que cada primo cuadrático divide a un único primo racional p (el que divide a su norma) y que $\mathrm{N}(\pi)$ tiene que ser p o p^2. Además, cada primo racional p se descompone a lo sumo en dos factores primos, que pueden ser asociados o no, lo que nos lleva a la adaptación obvia de la definición 4.5:

Definición 5.4 Si p es un primo racional, se dice que:

1. p se *ramifica* en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ si se descompone como $p = \pm\pi^2$, donde π es primo de norma p.

2. p se *escinde* en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ si se descompone en dos factores primos no asociados $p = \pi_1\pi_2$ (de norma p).

3. p se *conserva* (primo) en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ si es primo en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ (y entonces tiene norma p^2).

De nuevo no hay más posibilidades, pero lo que ya no es cierto es que los primos racionales que se escinden son los congruentes con 1 módulo 4. Esto nos lleva ya a estudiar la "personalidad propia" del anillo $\mathbb{Z}\big[\sqrt{-2}\,\big]$. Por ejemplo:

- $2 = -(\sqrt{-2})^2$, luego 2 se ramifica en $\mathbb{Z}\big[\sqrt{-2}\,\big]$.

- En cambio, $3 = 1^2 + 2 \cdot 1^2 = (1 + \sqrt{-2})(1 - \sqrt{-2})$ se escinde en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ (mientras que se conserva primo en $\mathbb{Z}[i]$).

- Por otra parte, 5 no puede descomponerse en la forma $x^2 + 2y^2$, luego se conserva primo en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ (pero se escinde en $\mathbb{Z}[i]$).

Ejercicio: Encontrar un primo que se escinda tanto en $\mathbb{Z}[i]$ como en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ y otro que se conserve en ambos anillos.

Si el lector se ha tomado en serio el problema de estudiar qué números son de la forma x^2+2y^2, ya debería tener formada una conjetura sobre qué primos se escinden en $\mathbb{Z}\big[\sqrt{-2}\,\big]$. Ahora observemos que, si bien superficialmente la escisión de primos en los anillos $\mathbb{Z}[i]$ y $\mathbb{Z}\big[\sqrt{-2}\,\big]$ parece obedecer a criterios completamente distintos, cuando se contempla la situación desde el nivel de abstracción adecuado se aprecia que la situación "es la misma". El teorema 4.6 es válido para $\mathbb{Z}\big[\sqrt{-2}\,\big]$ con las mínimas adaptaciones obvias y con la misma demostración:

Teorema 5.5 *Sea p un primo racional. Entonces:*

1. *Si $x^2+\bar{2}=(x-\bar{c})^2$ en $\mathbb{Z}_p[x]$, entonces $p=\epsilon\pi^2$, donde $\pi=(p,\sqrt{-2}-c)$ es un primo cuadrático que cumple $\sqrt{-2}\equiv c \pmod{\pi}$.*

2. *Si $x^2+\bar{2}=(x-\bar{c}_1)(x-\bar{c}_2)$ en $\mathbb{Z}_p[x]$, donde $c_1\not\equiv c_2 \pmod{p}$, entonces $p=\pi_1\pi_2$, donde $\pi_i=(p,\sqrt{-2}-c_i)$ son primos cuadráticos no asociados tales que $\sqrt{-2}\equiv c_j \pmod{\pi_j}$.*

3. *Si $x^2+\bar{2}$ es irreducible en $\mathbb{Z}_p[x]$, entonces p se conserva primo y $\sqrt{-2}$ no es congruente módulo p con ningún entero racional.*

No vamos a repetir la prueba, pues no consiste sino en repetir palabra por palabra la dada para los enteros de Gauss sin más que cambiar i por $\sqrt{-2}$ y x^2+1 por x^2+2, pero vamos a mostrar dos ejemplos que ilustren la situación en este caso:

Como 5 se conserva primo, tenemos que k_5 es un cuerpo de 25 elementos, formado por las clases $\bar{a}+\bar{b}\overline{\sqrt{-2}}$, con $\bar{a}$ y $\bar{b}$ en $\mathbb{Z}_5$. En realidad es razonable escribir $\sqrt{-2}$ en lugar de $\overline{\sqrt{-2}}$, ya que esta clase es una raíz cuadrada de -2 en k_5. Los elementos de este cuerpo se operan así:

$$(\bar{3}+\bar{2}\ \sqrt{-2})+(\bar{2}+\bar{4}\ \sqrt{-2})=\sqrt{-2},$$

$$(\bar{3}+\bar{2}\ \sqrt{-2})(\bar{2}+\bar{4}\ \sqrt{-2})=\bar{3}\cdot\bar{2}-\bar{2}\cdot\bar{2}\cdot\bar{4}+(\bar{3}\cdot\bar{4}+\bar{2}\cdot\bar{2})\sqrt{-2}=\bar{2}-\sqrt{-2}.$$

La prueba del teorema se basa en que el hecho de que 5 sea primo en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ implica que $\sqrt{-2}$ no puede pertenecer a $\mathbb{Z}_5$ y, por lo tanto, el polinomio x^2+2 no puede tener raíces en $\mathbb{Z}_5$.

Por el contrario, si consideramos $p=17=(3+2\sqrt{-2})(3-2\sqrt{-2})$ y tomamos $\pi=3+2\sqrt{-2}$, sucede que el cuerpo k_π es $\mathbb{Z}_{17}$, es decir, que todo entero de $\mathbb{Z}\big[\sqrt{-2}\,\big]$ es congruente con un entero racional módulo π. Ello se debe a que $\sqrt{-2}$ es congruente con un entero racional módulo π. Esto se deduce de la ecuación

$$\bar{3}+\bar{2}\sqrt{-2}=\bar{0},$$

de la que podemos despejar $\sqrt{-2}$. Para ello necesitamos el inverso de 2 módulo 17. Como $\bar{8}\cdot\bar{2} = -\bar{1}$, es claro que $-\bar{8}$ es el inverso que buscamos, luego $\bar{3} + \bar{2}\overline{\sqrt{-2}} = \bar{0}$ implica que $\bar{2}\overline{\sqrt{-2}} = -\bar{3}$, luego $\overline{\sqrt{-2}} = -\bar{8}(-\bar{3}) = \bar{6}$, y así:

$$\sqrt{-2} \equiv 6 \text{ (mód } \pi).$$

Alternativamente, podíamos haber (casi) predicho el valor 6 observando que

$$x^2 + 2 \equiv (x+6)(x-6) \text{ (mód } \pi).$$

Lo que muestra la prueba del teorema 5.5 es que, como $\sqrt{-2}$ es raíz de x^2+2 en k_π, si éste polinomio tiene ya sus raíces en $\mathbb{Z}_5$, entonces $\sqrt{-2}$ tiene que ser una de ellas. En este caso hemos visto que π hace que $\sqrt{-2}$ sea congruente con 6. Si hubiéramos tomado el primo $\bar{\pi} = 3 - 2\sqrt{-2}$, habríamos llegado al -6. ∎

Ahora sabemos que el comportamiento de un primo racional p en $\mathbb{Z}[\sqrt{-2}]$ se corresponde con el número de raíces distintas que tiene el polinomio x^2+2 módulo p, lo cual a su vez, si p es impar, depende únicamente de si su discriminante $\Delta = -8$ es 0 (en cuyo caso hay una única raíz y p se ramifica), es no nulo y es un resto cuadrático módulo p (en cuyo caso hay dos raíces y p se escinde) o bien es no nulo y no es un resto cuadrático módulo p (en cuyo caso no hay raíces y p se conserva primo).

Puesto que $\Delta = -8$ sólo es nulo módulo 2, concluimos que:

$$2 = -(\sqrt{-2})^2 \textit{ es el único primo racional que se ramifica en } \mathbb{Z}[\sqrt{-2}].$$

Falta determinar qué primos impares se escinden y cuáles se conservan. Sabemos que esto depende únicamente de si $\Delta = -8$ es un resto cuadrático o no cuadrático módulo p, lo que claramente equivale a que lo sea -2. Sin embargo, el -8 es relevante, porque es el valor adecuado para preguntarnos si se dará en este caso el mismo tipo de "reciprocidad" que nos hemos encontrado al analizar la aritmética de los enteros de Gauss: allí vimos que el carácter de un primo impar p (si se escinde o ramifica) depende en principio del carácter cuadrático de -4 (o -1) módulo p, y demostramos que también depende del resto de p módulo -4 (o 4).

Ahora podemos plantearnos si el carácter de un primo impar p en $\mathbb{Z}[\sqrt{-2}]$, que en principio depende del resto de -8 módulo p, también resulta depender del resto de p módulo -8 (u 8), y la respuesta ¡también es afirmativa! ¡Se da la misma clase de "reciprocidad" en este caso!

Si sacamos de la tabla dada al principio del capítulo la lista de los primos impares que son de la forma $x^2 + 2y^2$, es decir, los primos que se escinden en $\mathbb{Z}[\sqrt{-2}]$, resultan ser:

$$3,\quad 11,\quad 17,\quad 19,\quad 41,\quad 43,\quad 59,\quad 67,\quad 73,\quad 83,\quad 97,\quad \ldots$$

y vemos que son precisamente los que cumplen $p \equiv 1, 3$ (mód 8). Por lo tanto, nuestra conjetura es:

Teorema 5.6 *Si p es un primo racional entonces:*

- *p se ramifica en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ si y sólo si $p = 2$.*
- *p se escinde en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ si y sólo si $p \equiv 1, 3$ (mód 8).*
- *p se conserva en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ si y sólo si $p \equiv 5, 7$ (mód 8).*

El núcleo de este capítulo estará dedicado a demostrar esta conjetura, pero ahora vamos a ver cómo resuelve el problema de los números expresables en la forma $n = x^2 + 2y^2$.

Recordemos únicamente que hemos visto que para probar el teorema anterior basta probar que -2 es un resto cuadrático módulo p si y sólo si $p \equiv 1, 3$ (mód 8).

Ejercicio: Adaptar la prueba del teorema 4.8 para demostrar que el número de clases de congruencia módulo α en $\mathbb{Z}[\sqrt{-2}]$ es $\mathrm{N}(\alpha)$.

Ejercicio: Mostrar representantes de todas las clases de congruencia módulo 3 en $\mathbb{Z}[\sqrt{-2}]$. Determinar cuáles corresponden a unidades.

Números de la forma x^2+2y^2 El lector que haya asimilado bien la solución al problema de los números que pueden expresarse como suma de dos cuadrados debería ser capaz de enunciar una conjetura para el caso de los números de la forma $x^2 + 2y^2$ y demostrarla siguiendo exactamente los mismos pasos que seguimos en el capítulo anterior.

> *Un número natural n es de la forma $n = x^2 + 2y^2$ si y sólo si los primos que lo dividen con exponente impar son de la forma $p = 2$ o bien $p \equiv 1, 3$ (mód 8).*

Para probarlo observamos que n es de la forma x^2+2y^2 si y sólo si $n = \mathrm{N}(\alpha)$, para cierto entero cuadrático α de $\mathbb{Z}\big[\sqrt{-2}\,\big]$. Un tal α puede descomponerse en primos cuadráticos en la forma

$$\pm\lambda^{e_0} p_1^{e_1} \cdots p_k^{e_k} \pi_{k+1}^{e_{k+1}} \cdots \pi_r^{e_r},$$

donde hemos distinguido el primo $\lambda = \sqrt{-2}$ de norma 2, los primos p_j de norma p_j^2 y los primos π_j de norma prima p_j. Al calcular la norma queda

$$n = \mathrm{N}(\alpha) = 2^{e_0} p_1^{2e_1} \cdots p_k^{2e_k} p_{k+1}^{e_{k+1}} \cdots p_r^{e_r},$$

donde $p_1, \ldots, p_k$ son primos que se conservan (es decir, primos $p_j \equiv 5, 7$ (mód 8)) y $p_{k+1}, \ldots, p_r$ son primos que se ramifican.

Vemos que la única restricción para que un número natural n sea de esta forma, es decir, que sea la norma de un entero de Gauss, es que los primos $p \equiv 5, 7$ (mód 8) que dividan a n tienen que aparecer con exponente par. Equivalentemente, que todos los primos p que dividan a n con exponente impar tienen que ser $p = 2$ o bien $p \equiv 1, 3$ (mód 8). ∎

Ejemplo *Las únicas soluciones enteras de* $y^2 = x^3 - 2$ *son* $(x, y) = (3, \pm 5)$.

DEMOSTRACIÓN: Si x es par, entonces $y^2 \equiv -2 \pmod 8$, pero -2 no es un cuadrado módulo 8, por lo que x es impar, al igual que y. Expresamos la ecuación en la forma

$$x^3 = (y + \sqrt{-2})(y - \sqrt{-2}).$$

Los dos factores son primos entre sí en $\mathbb{Z}\big[\sqrt{-2}\,\big]$, pues un divisor primo común dividiría a su diferencia $2\sqrt{-2}$, luego tiene que ser $\sqrt{-2}$, que dividiría también a x, y entonces x sería par. Como las unidades de $\mathbb{Z}\big[\sqrt{-2}\,\big]$ son cubos, concluimos que cada factor es un cubo. Así:

$$y + \sqrt{-2} = (m + n\sqrt{-2})^3,$$

para ciertos enteros racionales m y n. Desarrollando:

$$y = m^3 - 6mn^2 = m(m^2 - 6n^2), \qquad 1 = 3m^2n - 2n^3 = n(3m^2 - 2n^2).$$

La segunda ecuación nos da que $n = \pm 1$. Si $n = -1$, entonces $1 = -(3m^2 - 2)$, luego $3m^2 = 1$, que no tiene solución entera. Por lo tanto, $n = 1 = 3m^2 - 2$, con lo que $m = \pm 1$. La primera ecuación nos da que $y = \pm 5$, luego $x = 3$. ∎

5.2 El símbolo de Legendre

Tenemos pendiente demostrar que -2 es un resto cuadrático módulo un primo p si y sólo si $p \equiv 1, 3 \pmod 8$ y, ya de paso encontraremos un criterio similar para el 2, que el lector no debería tener dificultad en conjeturar. El lector perezoso tiene aquí la lista de los primeros primos para los que 2 es un resto cuadrático:

$$7, \quad 17, \quad 23, \quad 31, \quad 41, \quad 47, \quad 71, \quad 73, \quad 79, \quad 89, \quad 97.$$

Sin embargo, antes de entrar en ello, vamos a introducir un formalismo útil para tratar con restos cuadráticos, que entre otras cosas nos mostrará que el problema para 2 es esencialmente el mismo que para -2, aunque las respuestas sean distintas.

Definición 5.7 Si $p>0$ es un primo impar y a es un número entero arbitrario, definimos el *símbolo de Legendre*

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{si } p \nmid a \text{ y } a \text{ es un resto cuadrático módulo } p, \\ 0 & \text{si } p \mid a, \\ -1 & \text{si } p \nmid a \text{ y } a \text{ es un resto no cuadrático módulo } p. \end{cases}$$

Por ejemplo, el teorema de Euler 3.38 se enuncia así en términos del símbolo de Legendre:

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}.$$

(Notemos que $(p-1)2$ es par si y sólo si $p \equiv 1 \pmod 4$.) En realidad se cumple algo más general:

Teorema 5.8 (Criterio de Euler) *Si $p > 0$ es un primo impar y a es un número entero, entonces*

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \text{ (mód } p).$$

DEMOSTRACIÓN: Si $p \mid a$, la conclusión es inmediata. Supongamos, pues que $p \nmid a$. Sea $b = a^{(p-1)/2}$. Entonces, por el teorema de Fermat 3.24, tenemos que $b^2 \equiv a^{p-1} \equiv 1$ (mód p), luego b es raíz del polinomio $x^2 - 1$, cuyas únicas raíces son ± 1, luego hemos probado que $a^{(p-1)/2} \equiv \pm 1$ (mód p).

Ahora basta aplicar 3.37 para $m = p$ y $n = 2$, con lo que $d = (2, p-1) = 2$. La conclusión es que $(a/p) = 1$ si y sólo si $a^{(p-1)/2} \equiv 1$ (mód p), luego también $(a/p) = -1$ si y sólo si $a^{(p-1)/2} \equiv -1$ (mód p), y así tenemos la congruencia del enunciado. ∎

Por ejemplo, si queremos saber si 3 es un resto cuadrático módulo 11, en lugar de ir calculando cuadrados a ver si alguno es 3, es más corto calcular

$$3^{(11-1)/2} = 3^5 \equiv 9 \cdot 9 \cdot 3 \equiv (-2)(-2) \cdot 3 = 12 \equiv 1 \text{ (mód } 11),$$

y así concluimos que 3 es un resto cuadrático módulo 11.

Una consecuencia notable del criterio de Euler es la siguiente:

Teorema 5.9 *Si $p > 0$ es un primo impar y a y b son enteros arbitrarios, entonces*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

DEMOSTRACIÓN: Por el criterio de Euler, ambos miembros son congruentes módulo p y, como ambos valen ± 1 y no es cierto que $1 \equiv -1$ (mód p), tiene que darse la igualdad. ∎

Esto implica que el cálculo de (a/p) puede reducirse al cálculo de símbolos de Legendre de la forma $(-1/p)$ (que ya sabemos calcular) y símbolos (q/p), donde $q > 0$ es primo.

Por ejemplo, de los teoremas 3.38 y 3.39 ahora podemos deducir:

Teorema 5.10 *Si $p > 3$ es primo, entonces:*

$$\left(\frac{-3}{p}\right) = 1 \quad \textit{si y sólo si} \quad p \equiv 1 \text{ (mód } 3).$$

$$\left(\frac{3}{p}\right) = 1 \quad \textit{si y sólo si} \quad p \equiv \pm 1 \text{ (mód } 12).$$

DEMOSTRACIÓN: La primera afirmación es una mera reformulación de 3.39, mientras que la segunda es consecuencia de que

$$\left(\frac{3}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{-3}{p}\right).$$

Basta considerar los cuatro restos posibles de p módulo 12 para comprobar que el producto vale 1 exactamente en los dos casos indicados:

p (mód 12)	1	5	−5	−1
$(-1/p)$	1	1	−1	−1
$(-3/p)$	1	−1	1	−1
$(3/p)$	1	−1	−1	1

■

A estas alturas el lector ya debía haber conjeturado que 2 es un resto cuadrático módulo p si y sólo si $p \equiv \pm 1$ (mód 8). En términos del símbolo de Legendre, esto se puede expresar así:

Teorema 5.11 *Si p es un primo impar, entonces*

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8},$$

de modo que 2 es un resto cuadrático módulo p si y sólo si $p \equiv \pm 1$ (mód 8).

En efecto, la equivalencia se debe a que $(p^2-1)/8$ es par si y sólo si

$$2^4 \mid p^2 - 1 = (p+1)(p-1)$$

lo cual equivale a que 8 divida a uno de los dos factores, pues es imposible que 4 divida a ambos (si un número es múltiplo de 4, al sumarle 2 deja de serlo).

Admitiendo el teorema anterior, el carácter cuadrático de -2 se sigue de la igualdad

$$\left(\frac{-2}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{2}{p}\right),$$

pues ambos factores dependen del resto de p módulo 8, por lo que basta calcular la tabla siguiente:

p	1	3	5	7
$(-1/p)$	1	−1	1	−1
$(2/p)$	1	−1	−1	1
$(-2/p)$	1	1	−1	−1

Concluimos que -2 es un resto cuadrático módulo un primo impar p si y sólo si $p \equiv 1, 3$ (mód 8), que es justo lo que necesitamos para completar la prueba del teorema 5.6. Así pues, este teorema quedará probado en cuanto hayamos probado 5.11. Nos ocupamos de ello en la sección siguiente.

5.3 El carácter cuadrático de 2

Euler conjeturó que 2 es un resto cuadrático módulo un primo impar $p > 0$ si y sólo si $p \equiv \pm 1$ (mód 8), pero sólo supo demostrar que si $p \equiv 1$ (mód 8), entonces 2 es un resto cuadrático módulo p, y ello dando por hecha la existencia de raíces primitivas, cosa que tampoco supo demostrar.

Existen muchas demostraciones sustancialmente distintas del teorema 5.11 y aquí vamos a dar una consistente en refinar el argumento de Euler mediante un poco de "ingeniería algebraica". Para entender a qué nos referimos empezaremos exponiendo el razonamiento de Euler.

Su motivación última podemos encontrarla en el álgebra de los números complejos. No vamos a necesitar este hecho aquí, pero si el lector está familiarizado con los números complejos sabrá que el número complejo de módulo 1 y argumento $\pi/4$, es decir,

$$\zeta = \cos 45^\circ + i \operatorname{sen} 45^\circ = \frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2},$$

es una raíz octava de la unidad (cumple $\zeta^8 = 1$) con la propiedad adicional de que 8 es el menor exponente no nulo que cumple esta relación. En particular, $\zeta^4 = -1$ (en principio, de $\zeta^8 = 1$ se deduce que $\zeta^4 = \pm 1$, pero de hecho se cumple que $\zeta^4 = -1$).

Esta fórmula muestra que a partir de una raíz cuadrada de 2 y de una raíz cuadrada de -1 podemos construir un ζ que cumple $\zeta^4 = -1$ y, recíprocamente, a partir de ζ es posible construir una raíz cuadrada de 2 y una raíz cuadrada de -1.

En efecto, la raíz cuadrada de -1 es ζ^2, mientras que $\sqrt{2}$ puede obtenerse como $\zeta + \bar{\zeta} = \sqrt{2}$. Aquí la barra indica la conjugación compleja, pero podemos sustituirla por algo "más universal" si observamos que, para números complejos de módulo 1, se cumple que $\bar{\zeta} = \zeta^{-1}$, por lo que la igualdad anterior equivale a $\zeta + \zeta^{-1} = \sqrt{2}$.

En estos términos, sucede que estas relaciones son válidas en cualquier cuerpo, no necesariamente el cuerpo de los números complejos. Si en un cuerpo k (de característica distinta de 2) tenemos una raíz cuadrada de -1, digamos i, y una raíz cuadrada de 2, digamos $\sqrt{2}$, entonces

$$\zeta = \frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}$$

cumple $\zeta^2 = i$, luego $\zeta^4 = -1$.

Recíprocamente —y esto es lo único que vamos a necesitar—, si en un cuerpo k (de característica distinta de 2) tenemos un elemento ζ que cumple $\zeta^4 = -1$, entonces $\zeta^2 = -\zeta^{-2}$, luego

$$(\zeta + \zeta^{-1})^2 = \zeta^2 + \zeta^{-2} + 2 = 2.$$

Pero Euler no razonó "en un cuerpo cualquiera", sino precisamente en $\mathbb{Z}_p$. Él argumentó que si $p \equiv 1$ (mód 8), entonces $8 \mid p-1$, luego si g es una raíz primitiva módulo p, entonces $g^{(p-1)/2} = -1$ (pues no es 1, pero su cuadrado es 1), luego $\zeta = g^{(p-1)/8}$ cumple que $\zeta^4 = -1$, luego, según acabamos de comprobar, $(\zeta+\zeta^{-1})^2 = 2$, lo que prueba que 2 es un resto cuadrático módulo p.

Observemos que para que $\mathbb{Z}_p$ contenga un ζ tal que $\zeta^4 = -1$ es necesario que $p \equiv 1$ (mód 8), pues, por el teorema de Fermat, $o_p(\zeta) = 8 \mid p-1$. Por lo

tanto, el argumento de Euler no da más de sí... ¡mientras nos mantengamos en $\mathbb{Z}_p$! La prueba que vamos a dar del teorema 5.11 consiste en observar que el razonamiento de Euler ¡no requiere que ζ esté en $\mathbb{Z}_p$!, sino que podemos añadir una raíz octava de la unidad a $\mathbb{Z}_p$ con la misma técnica de "ingeniería algebraica" con la que hemos añadido a $\mathbb{Q}$ una raíz cuadrada de -1 o de -2, es decir, mediante el teorema 4.1.

El teorema 4.1 requiere partir de un polinomio irreducible, pero x^4+1 no es necesariamente irreducible en $\mathbb{Z}_p[x]$. Por ejemplo, si $p=7$ tenemos que

$$x^4+\bar{1} = (x^2+\bar{3}x+\bar{1})(x^2-\bar{3}x+\bar{1})$$

y para $p=17$ tenemos que

$$x^4+\bar{1} = (x+\bar{2})(x-\bar{2})(x+\bar{8})(x-\bar{8}).$$

En cualquier caso, podemos tomar un factor irreducible $q(x) \mid x^4+1$ y considerar el anillo de restos k de $\mathbb{Z}_p[x]$ módulo $q(x)$, que es un cuerpo por 4.1. Como tiene obviamente característica p (porque $\mathbb{Z}_p[x]$ tiene característica p), podemos considerar que contiene a $\mathbb{Z}_p$ (identificando sus elementos con sus clases de restos módulo $q(x)$), pero además, contiene a $\zeta = [x]$, que cumple $q(\zeta)=0$.

Por ejemplo, si $p=7$ y $q(x)=x^2+\bar{3}x+\bar{1}$, se cumple que

$$[x^2+\bar{3}x+\bar{1}] = [x]^2+[\bar{3}][x]+[\bar{1}] = 0,$$

que es lo mismo que $q(\zeta)=0$.

Además, como $q(x)$ divide x^4+1, también se cumple que $\zeta^4+1=0$. En definitiva, acabamos de ver que existe un cuerpo k que contiene a $\mathbb{Z}_p$, así como a una raíz octava de la unidad ζ, que cumple $\zeta^4=-1$. Que ζ no esté necesariamente en $\mathbb{Z}_p$ es irrelevante.

Ahora llamamos $\sqrt{2} = \zeta+\zeta^{-1}$, y hemos visto que es realmente una raíz cuadrada de 2, es decir, una raíz del polinomio $x^2-\bar{2}$ que en principio está en k. Así, 2 es un resto cuadrático módulo p si y sólo si este polinomio tiene sus (dos) raíces en $\mathbb{Z}_p$, pero como no puede tener más de dos raíces en k y éstas son $\pm\sqrt{2}$, concluimos que 2 es un resto cuadrático módulo p si y sólo si $\sqrt{2}\in\mathbb{Z}_p$.

Por otra parte, hay un criterio muy simple para determinar si un elemento de k está o no en $\mathbb{Z}_p$. Por el teorema de Fermat, los elementos no nulos de $\mathbb{Z}_p$ cumplen $a^{p-1}=1$, luego $a^p=a$, pero esto también lo cumple el 0, luego los elementos de $\mathbb{Z}_p$ son raíces del polinomio x^p-x. Ahora bien, como este polinomio sólo puede tener p raíces, concluimos que un elemento a de k está en $\mathbb{Z}_p$ si y sólo si cumple $a^p=a$. Aplicándolo a $\sqrt{2}$, tenemos que está en $\mathbb{Z}_p$ si y sólo si

$$(\zeta+\zeta^{-1})^p = \zeta^p+\zeta^{-p} = \zeta+\zeta^{-1}.$$

Las potencias de ζ sólo dependen del resto de p módulo 8, y hay cuatro posibilidades:

1. Si $p \equiv 1 \pmod 8$, entonces $\zeta^p + \zeta^{-p} = \zeta + \zeta^{-1}$.
2. Si $p \equiv -1 \pmod 8$, entonces $\zeta^p + \zeta^{-p} = \zeta^{-1} + \zeta$.
3. Si $p \equiv 3 \pmod 8$, entonces $\zeta^p + \zeta^{-p} = \zeta^3 + \zeta^{-3} = -\zeta^{-1} - \zeta$ (pues la relación $\zeta^4 = -1$ implica que $\zeta^3 = -\zeta^{-1}$).
4. Si $p \equiv -3 \pmod 8$, entonces $\zeta^p + \zeta^{-p} = \zeta^{-3} + \zeta^3 = -\zeta - \zeta^{-1}$.

En resumen:

$$(\sqrt{2})^p = \begin{cases} \sqrt{2} & \text{si } p \equiv \pm 1 \pmod 8, \\ -\sqrt{2} & \text{si } p \equiv \pm 3 \pmod 8. \end{cases}$$

y la conclusión es que $(2/p) = 1$ si y sólo si $p \equiv \pm 1 \pmod 8$, como queríamos probar. ∎

Ejercicio: Probar que si p es un primo impar, el polinomio $x^4 + 1$ no es irreducible en $\mathbb{Z}_p[x]$. AYUDA: Si ζ es una raíz en un cuerpo que contenga a $\mathbb{Z}_p$ y $p \not\equiv 1 \pmod 8$, entonces un factor irreducible es $(x - \zeta)(x - \zeta^p) \in \mathbb{Z}_p[x]$. Más explícitamente:

$$x^4 + 1 = \begin{cases} (x^2 + \sqrt{-2}x - 1)(x^2 - \sqrt{-2}x - 1) & \text{si } p \equiv 3 \pmod 8, \\ (x^2 + \sqrt{-1})(x^2 - \sqrt{-1}) & \text{si } p \equiv 5 \pmod 8, \\ (x^2 + \sqrt{2}x + 1)(x^2 - \sqrt{2}x + 1) & \text{si } p \equiv 7 \pmod 8. \end{cases}$$

Encontrar los factores irreducibles para $p = 3, 5, 7$.

Ejercicio: Demostrar que $(-1/p) = (-1)^{(p-1)/2}$ con un razonamiento análogo al anterior, usando el polinomio $x^2 + 1$ en lugar de $x^4 + 1$.

Veamos algunas aplicaciones:

Ejemplo *La ecuación $y^2 = x^3 - 6$ no tiene soluciones enteras.*

DEMOSTRACIÓN: Si x es par, entonces $y^2 \equiv 2 \pmod 8$, pero 2 no es un cuadrado módulo 8. Por lo tanto x es impar, luego y también es impar, luego $y^2 \equiv 1 \pmod 8$, y a su vez $x \equiv x^3 \equiv -1 \pmod 8$ (la primera congruencia la cumple todo x impar). Escribimos la ecuación como

$$y^2 - 2 = x^8 - 8 = (x - 2)(x^2 + 2x + 4).$$

Tenemos que $x^2 + 2x + 4 = (x + 1)^2 + 3 > 0$ y $x^2 + 2x + 4 \equiv 3 \pmod 8$, luego es divisible entre un primo $p \equiv \pm 3 \pmod 8$, ya que si todos los divisores primos fueran congruentes con $\pm 1 \pmod 8$, lo mismo le sucedería a su producto. Como $p \mid y^2 - 2$, resulta que $2 \equiv y^2 \pmod p$, pero $(2/p) = -1$, y tenemos una contradicción. ∎

Ejemplo *La ecuación $y^2 = x^3 + 6$ no tiene soluciones enteras.*

Como en el ejemplo anterior concluimos que x e y son impares, y que además $x \equiv 3 \pmod 8$. Expresamos la ecuación como

$$y^2 + 2 = x^3 + 8 = (x + 2)(x^2 - 2x + 4),$$

donde $x^2 - 2x + 4 \equiv -1 \pmod 8$. Si $p > 0$ divide a este número, entonces $-2 \equiv y^2 \pmod p$, es decir, $(-2/p) = 1$, luego $p \equiv 1, 3 \pmod 8$, pero el producto de números congruentes con $1, 3$ móduo 8 cumple esta misma propiedad y, como $x^2 - 2x + 4 = (x-1)^2 + 3 > 0$, tiene que ser congruente con $1, 3$ módulo 8, cuando también hemos visto que tiene que ser $\equiv -1 \pmod 8$. ∎

Ejemplo *La ecuación $y^2 = x^3 - 24$ no tiene soluciones enteras.*

DEMOSTRACIÓN: Escribamos la ecuación en la forma

$$y^2 + 16 = x^3 - 8 = (x-2)(x^2 + 2x + 4),$$

donde $x^2 + 2x + 4 = (x+1)^2 + 3 \geq 3$. Si x es impar, esta expresión muestra que $x^2 + 2x + 4 \equiv -1 \pmod 4$, luego este número es divisible entre un primo de la forma $p \equiv -1 \pmod 4$. Pero entonces $-16 \equiv y^2 \pmod p$ implica que $(-1/4) = 1$, y tenemos una contradicción. Por lo tanto x es par, al igual que y. La forma original de la ecuación muestra que, de hecho, $8 \mid y$, luego podemos expresar $x = 2x'$, $y = 4y'$, y la ecuación se convierte en

$$2y'^2 = x'^3 - 3.$$

Esto implica que $x' \geq 3$ es impar. Como $y'^2 \equiv 0, 1, 4 \pmod 8$, tenemos que $x' \equiv x'^3 \equiv \pm 3 \pmod 8$. Expresamos la ecuación como

$$2(y'^2 + 2) = x'^3 + 1 = (x'+1)(x'^2 - x' + 1),$$

donde $x'^2 - x' + 1 \equiv -1, -3 \pmod 8$. Si $p > 0$ es un divisor primo de este número, entonces $y'^2 \equiv -2 \pmod p$, luego se cumple que $p \equiv 1, 3 \pmod 8$. Como $x'^2 - x' + 1 = (x' - 1/2)^2 + 3/4 > 0$, también $x'^2 - x' + 1 \equiv 1, 3 \pmod 8$, y tenemos una contradicción. ∎

Ejemplo: Números de Fermat En la sección 3.5 estudiamos los primos de Mersenne, que son los primos de la forma $2^n - 1$ (y vimos que una condición necesaria para que un número de esta forma sea primo es que el exponente n lo sea también). Ahora vamos a estudiar los llamados *primos de Fermat,* que son los primos de la forma $2^n + 1$. La tabla siguiente muestra que en este caso el exponente no necesita ser primo:

n	**1**	**2**	3	**4**	5	6	7	**8**	9	10
$2^n + 1$	3	5	9	17	33	65	129	257	513	1 025

Por el contrario, resultan primos los números correspondientes a los valores $n = 1, 2, 4, 8, \ldots$ Fermat conjeturó basándose en esto que los primos que hoy llevan su nombre son exactamente los números de la forma $2^{2^n} + 1$.

Ciertamente todo primo de Fermat es de esta forma: sea $p = 2^n + 1$ un primo de Fermat, entonces $2^n \equiv -1 \pmod p$, luego $2^{2n} \equiv 1 \pmod p$. Así, por el teorema de Fermat, $o_p(2) \mid p - 1 = 2^n$, luego $o_p(2) = 2^e$, para cierto

$1 \le e \le n$. Pero, por definición de orden, $o_p(2) = 2^e \mid 2n$, luego $2^{e-1} \mid n$. Tenemos que $\overline{2^{2^{e-1}}} \neq \bar{1}$, pero el cuadrado de esta clase es $\bar{1}$, luego tiene que ser $\overline{2^{2^{e-1}}} = -\bar{1}$. Así pues, $p = 2^n + 1 \mid 2^{2^{e-1}} + 1 \le 2^n + 1 = p$, luego $p = 2^{2^{e-1}} + 1$, es decir, $n = 2^{e-1}$.

Queda por ver si todo número de la forma $F_n = 2^{2^n} + 1$ es primo. Los números de esta forma se llaman *números de Fermat*. Volvamos a la tabla y ampliémosla:

n	0	1	2	3	4	5
F_n	3	5	17	257	65 537	4 294 967 297

No es difícil comprobar que 257 es primo, pero, sin la ayuda de un ordenador, comprobar si lo es 65 537 es ya bastante laborioso. Vamos a ver que podemos determinarlo fácilmente sin necesidad siquiera de una calculadora, mediante técnicas similares a las que empleamos en el estudio de los primos de Mersenne.

Para ello observamos que si $p \mid F_n$ es primo, entonces $2^{2^n} \equiv -1 \pmod p$, luego $2^{2^{n+1}} \equiv 1 \pmod p$, luego $o_p(2) \mid 2^{n+1}$, pero $o_p(2) \nmid 2^n$, o sería $2^{2^n} \equiv 1 \pmod p$, luego $o_p(2) = 2^{n+1}$.

Por otra parte, $o_p(2) \mid p - 1$, luego $p = 2^{n+1}r + 1$. En particular, si $n \ge 2$ se cumple que $p \equiv 1 \pmod 8$, luego $(2/p) = 1$, luego existe un entero x tal que $2 \equiv x^2 \pmod p$. Es fácil ver entonces que $o_p(x) = 2^{n+2}$, luego $p = 2^{n+2}k + 1$.

Así pues:

> *Si $n \ge 2$, todo divisor primo de $F_n = 2^{2^n} + 1$ tiene que ser de la forma $p = 2^{n+2}k + 1$.*

Por ejemplo, si $F_3 = 257$ no fuera primo, como $256 < 17^2$, tendría un divisor primo menor que 17, pero a la vez tendría que ser de la forma $p = 32k+1$, luego concluimos que 256 es primo (sin necesidad de cálculo alguno). Similarmente, si F_4 no fuera primo, tendría un divisor primo de la forma $p = 64k + 1 < 257$. Los primeros valores de la sucesión $64k + 1$ son:

$$65,\ 129,\ 193,\ 257,\ \ldots$$

y el único primo en ella menor que 257 es 193. Por lo tanto, F_4 será primo si y sólo si no es divisible entre 193. Comprobar que esto es cierto no supone ninguna dificultad, luego, en efecto, $F_4 = 65\,537$ es primo.

El siguiente posible primo de Fermat es muchísimo mayor. Si aplicamos el mismo razonamiento concluimos que si F_5 no es primo, entonces tiene un divisor primo de la forma $p = 128k + 1$ y $p < 65\,536$. Esto nos deja todavía un gran número de candidatos. Los diez primeros son

$$129,\ \mathbf{257},\ 385,\ 513,\ \mathbf{641},\ \mathbf{769},\ 897,\ 1\,025,\ \mathbf{1\,153},\ 1\,281,\ \ldots$$

(los primos están en negrita).

El primer primo es fácil de descartar por su forma, ya que se trata de 2^8+1 y módulo 2^8+1 se cumple $\bar{2}^{2^5}+\bar{1}=(\bar{2}^8)^4+\bar{1}=-\bar{1}^4+\bar{1}=\bar{2}\neq\bar{0}$.

Respecto al siguiente, las potencias de 2 módulo 641 son las siguientes:

n	1	2	4	8	16	32
2^n	2	4	16	256	154	-1

Así pues, $\bar{2}^{2^5}+\bar{1}=\bar{0}$, es decir, $641 \mid 2^{2^5}+1$ que no es, por tanto, primo. Euler anunció este hecho en 1732. Además, un cálculo rutinario nos da que[1]

$$F_5 = 641 \cdot 6\,700\,417$$

No se sabe si Euler se paró a comprobar que $p = 6\,700\,471$ es un número primo, pero es probable que lo hiciera, porque tenía casi todo el trabajo hecho. Sólo hay 4 primos de la forma $128k+1$ menores que su raíz cuadrada (que es 2588.52) sin contar 257, que ya lo hemos descartado. A saber:

$$641, \quad 769, \quad 1153, \quad 1409.$$

Basta comprobar que p no es divisible por ninguno de ellos para concluir que es primo.[2] Hoy en día no se conoce ningún otro primo de Fermat distinto de los cinco que ya hemos encontrado:

$$3, \quad 5, \quad 17, \quad 257, \quad 65\,537.$$

Ejercicio: Demostrar que todos los números de Fermat $F_n = 2^{2^n}+1$ (sean primos o no) para $n \geq 2$ terminan en 7.

Primos de Sophie Germain y primos de Mersenne Se dice que un primo p es un *primo de Sophie Germain* si $2p+1$ también es primo.

Los primeros primos de Sophie Germain son:

$$2, 3, 5, 11, 23, 29, 41, 53, 83, 89, \ldots$$

Se conjetura que hay infinitos. El resultado siguiente es de Euler:

> *Si p es un primo de Sophie Germain y $p \equiv -1$ (mód 4), entonces el número de Mersenne $M_p = 2^p - 1$ cumple que $2p+1 \mid M_p$, por lo que no es primo, salvo en el caso $p = 3$, en el que $2p+1 = 7 = M_3$.*

[1] Un cálculo alternativo se basa en observar que $641 = 2^7 \cdot 5 + 1 = 2^4 + 5^4$. Esto hace que $2^7 \cdot 5 \equiv -1$ (mód 641), luego $2^{28} \cdot 5^4 \equiv 1$ (mód 641) y $5^4 \equiv -2^4$ (mód 641). Por lo tanto, $2^{32} \equiv -1$ (mód 641).

[2] En realidad Euler habría tenido que trabajar un poco más, pues él consideró primos de la forma $64k+1$. La reducción a $128k+1$ se debe a Lucas.

Los primeros valores de p a los que es aplicable este criterio son

$$p = 3, 11, 23, 83, \ldots$$

En particular tenemos una prueba de que $23 \mid M_{11} = 2\,047$ mucho más simple que la que obtuvimos en la sección 3.5, y también podemos asegurar, por ejemplo, que $47 \mid M_{23} = 8\,388\,609$.

DEMOSTRACIÓN: Sea $q = 2p + 1$. Por el teorema de Fermat, tenemos que $2^{2p} \equiv 1 \pmod{q}$, luego $2^p \equiv \pm 1 \pmod{q}$ (pues en el cuerpo $\mathbb{Z}_q$ sólo puede haber dos raíces cuadradas de 1, que son ± 1). Basta probar que no puede darse el caso $2^p \equiv -1 \pmod{q}$. Si así fuera, tendríamos que

$$2^{p+1} = (2^{(p+1)/2})^2 \equiv -2 \pmod{q},$$

luego $(-2/q) = 1$. Sin embargo, como $p = 4k - 1$, tenemos que

$$q = 2(4k - 1) + 1 = 8k - 1,$$

luego $(-2/q) = -1$ y tenemos una contradicción. ∎

Euler y M_{31} Euler encontró un criterio para descartar primos a la hora de factorizar números de Mersenne:

> *Si p y q son primos impares y $q \mid M_p = 2^p - 1$, entonces se cumple que $q \equiv \pm 1 \pmod{8}$.*

En efecto, tenemos que $2^p \equiv 1 \pmod{q}$, luego, por el teorema de Fermat, $o_q(2) = p \mid q - 1$. Pero, como p es impar, de hecho $p \mid (q-1)/2$, luego de hecho tenemos que $2^{(q-1)/2} \equiv 1 \pmod{q}$, y el criterio de Euler 5.8 nos da que $(2/q) = 1$, luego $q \equiv \pm 1 \pmod{8}$. ∎

Euler usó este hecho en 1772 para probar que el número de Mersenne

$$M_{31} = 2\,147\,483\,647$$

es primo, con lo que éste pasó a ser el mayor primo conocido, superando al primo 6 700 417, que él mismo había encontrado 40 años antes. Concretamente, observamos que si $q \mid M_{31}$, entonces

$$q \equiv 1 \pmod{31}, \qquad q \equiv \pm 1 \pmod{8}.$$

El teorema chino del resto nos da que

$$q \equiv 1, 63 \pmod{248}.$$

Esto reduce los divisores primos de M_{31} menores que $\sqrt{M_{31}}$ a un total de 42 primos congruentes con 1 (el mayor es $q = 45\,137$) y 43 primos congruentes con 63 (hasta $q = 46\,377$), es decir, a un total de 85 divisiones que Euler tuvo que realizar para comprobar que M_{31} es primo. ∎

En la sección 3.5 hemos probado que M_{13} es primo comprobando que no es divisible entre 53 y 79. Teniendo en cuenta el criterio de Euler, no era necesario comprobar el 53.

Fermat y M_{37} En la introducción señalamos que Fermat había comunicado a Mersenne en una carta que M_{37} no es primo. En la sección 3.5 hemos visto que un divisor primo de M_{37} tiene que ser de la forma $q = 74k+1$. Los primeros primos de esta forma son $q = 149, 223, \ldots$, y el primero lo podemos descartar por el criterio de Euler, ya que $149 \equiv 5 \pmod{8}$. Por lo tanto, al primer intento encontramos la descomposición que encontró Fermat:

$$137\,438\,953\,471 = 2^{37} - 1 = 223 \cdot 616\,318\,177.$$

Queda pendiente determinar si el segundo factor es primo. Si no lo es, sus factores primos dividen también a M_{37}, luego sabemos que son de la forma $q = 74k+1$ y $q \equiv \pm 1 \pmod{8}$. Hay 73 primos que cumplen la primera condición y que son menores que

$$\sqrt{616\,318\,177} = 28\,825.8\ldots$$

No obstante, si incorporamos la condición de Euler el número se reduce. Si $q \equiv 1 \pmod{8}$ entonces es de la forma $q \equiv 296k + 1$, mientras que si se cumple $q \equiv -1 \pmod{8}$ entonces $q \equiv 296k + 223$. Esto reduce los primos posibles a 32.

593	1 481	1 777	3 257	4 441	6 217	7 993	9 473	9 769
10 657	12 433	13 913	17 761	18 353	20 129	21 017	21 313	23 977
1 999	2 591	2 887	4 663	6 143	7 919	8 807	9 103	11 471
14 431	15 319	19 463	19 759	23 311				

Probablemente, Fermat comprobó los 73 primos uno por uno para concluir que los dos factores de la descomposición de M_{37} son primos, hecho que necesitaba para resolver el reto de Mersenne que hemos discutido en la introducción. ∎

5.4 La congruencia $y^2 \equiv ax^2 + bx + c \pmod{p}$

Teorema 5.12 *Sea p un primo impar y sean a, b, c números enteros tales que $(a, p) = 1$. Sea $D = b^2 - 4ac$. Entonces el número de soluciones de la congruencia*

$$y^2 \equiv ax^2 + bx + c \pmod{p},$$

es decir, el número de soluciones en $\mathbb{Z}_p \times \mathbb{Z}_p$ de la ecuación

$$y^2 = \bar{a}x^2 + \bar{b}x + \bar{c}$$

es

$$N = \begin{cases} p - \left(\frac{a}{p}\right) & \text{si } p \nmid D, \\ p + (p-1)\left(\frac{a}{p}\right) & \text{si } p \mid D. \end{cases}$$

Demostración: Para cada $x \in \mathbb{Z}_p$, la ecuación tendrá solución para dos valores de y, para uno o para ninguno según que $ax^2 + bx + c$ sea un resto

cuadrático (no nulo) módulo p, sea 0 o sea un resto no cuadrático módulo p. Equivalentemente, tendrá

$$1+\left(\frac{ax^2+bx+c}{p}\right)$$

soluciones. Por lo tanto, el número total de soluciones de la congruencia es

$$N=\sum_{x=0}^{p-1}\left(1+\left(\frac{ax^2+bx+c}{p}\right)\right)=p+\sum_{x=0}^{p-1}\left(\frac{ax^2+bx+c}{p}\right).$$

La relación explícita entre un polinomio de grado 2 y su discriminante es que

$$4a(ax^2+bx+c)=(2ax+b)^2-D,$$

luego

$$\left(\frac{a}{p}\right)\left(\frac{ax^2+bx+c}{p}\right)=\left(\frac{(2ax+b)^2-D}{p}\right),$$

o también

$$\left(\frac{ax^2+bx+c}{p}\right)=\left(\frac{a}{p}\right)\left(\frac{(2ax+b)^2-D}{p}\right).$$

Como $2a$ es una unidad en $\mathbb{Z}_p$, la aplicación $x\mapsto 2ax+b$ biyecta las clases de $\mathbb{Z}_p$ consigo mismas, luego

$$N=p+\left(\frac{a}{p}\right)\sum_{x=0}^{p-1}\left(\frac{(2ax+b)^2-D}{p}\right)=p+\left(\frac{a}{p}\right)\sum_{y=0}^{p-1}\left(\frac{y^2-D}{p}\right).$$

Llamemos

$$S(D)=\sum_{y=0}^{p-1}\left(\frac{y^2-D}{p}\right),$$

de modo que $N=p+\left(\frac{a}{p}\right)S(D)$.

Si $p\mid D$ tenemos que

$$S(D)=\sum_{y=0}^{p-1}\left(\frac{y}{p}\right)^2=p-1,$$

lo que nos da la fórmula del enunciado para este caso.

Supongamos que $p\nmid D$. Entonces, cuando y varía entre 0 y $p-1$, tenemos que y^2 recorre dos veces cada resto cuadrático no nulo y toma una vez el valor 0. Por lo tanto,

$$S(D)=\sum_{z=0}^{p-1}\left(\left(\frac{z}{p}\right)+1\right)\left(\frac{z-D}{p}\right).$$

En efecto, en esta suma, cuando $z=0$ contamos $(z-D/p)$ una vez, cuando z es un resto cuadrático no nulo lo contamos dos veces y cuando es un resto no cuadrático no lo contamos.

Consideremos ahora la aplicación $z \mapsto z^*$ definida sobre $\{0, 1, \ldots, p-1\}$ que cumple $0^* = 0$ y que a cada z no nulo le asigna el representante de la clase $\bar{z}^{-1}$. Así $(z/p) = (z^*/p)$, luego

$$S(D) = \sum_{z=0}^{p-1} \left(\left(\frac{z^*}{p}\right) + 1 \right) \left(\frac{z-D}{p}\right) = \sum_{z=0}^{p-1} \left(\frac{zz^* - Dz^*}{p}\right) + \sum_{z=0}^{p-1} \left(\frac{z-D}{p}\right)$$

$$= \sum_{z=1}^{p-1} \left(\frac{1-Dz^*}{p}\right) + \sum_{z=0}^{p-1} \left(\frac{z-D}{p}\right) = -1 + \sum_{z=0}^{p-1} \left(\frac{1-Dz^*}{p}\right) + \sum_{z=0}^{p-1} \left(\frac{z-D}{p}\right).$$

Ahora bien, cuando z varía entre 0 y $p-1$, las clases $\overline{1-Dz^*}$ y $\overline{z-D}$ recorren todo $\mathbb{Z}_p$, luego

$$S(D) = -1 + 2\sum_{w=1}^{p-1} \left(\frac{w}{p}\right),$$

pero la última suma es nula, porque la mitad de los elementos de U_p son restos cuadráticos y la otra mitad restos no cuadráticos, luego en la suma hay $(p-1)/2$ sumandos iguales a 1 y otros tantos iguales a -1. Concluimos que $S(D) = -1$.

∎

Ejercicio: Calcular las soluciones de la congruencia $y^2 \equiv 2x^2 + 3x + 3$ (mód 5) y comprobar explícitamente que su número es el dado por el teorema anterior.

Ejercicio: Calcular el número de veces que $x^2 + y^2$ toma un valor c en $\mathbb{Z}_p$.

Capítulo VI

Enteros de Eisenstein

En la sección 4.2 construimos los enteros de Gauss añadiéndole a $\mathbb{Q}$ una raíz cuadrada de -1 o, lo que es lo mismo, una raíz cuarta de 1, pero no una raíz cuarta cualquiera (pues 1 y -1 ya son raíces cuartas de 1), sino una raíz cuarta de orden precisamente 4, una que hay que elevar a 4 para conseguir el 1. En general, un elemento α de un anillo es una raíz *n-sima primitiva primitiva* de la unidad si n es el menor exponente no nulo que cumple $\alpha^n = 1$.

En este capítulo vamos a añadir a $\mathbb{Q}$ una raíz cúbica primitiva de la unidad, y veremos que mediante el anillo de enteros resultante podremos demostrar cosas como que la ecuación $x^3 + y^3 = z^3$ no tiene soluciones enteras no triviales (es decir, soluciones en las que $xyz \neq 0$), que es el caso más simple del Último Teorema de Fermat.

6.1 La aritmética de los enteros de Eisenstein

Una raíz cúbica de la unidad es una raíz ζ del polinomio

$$x^3 - 1 = (x-1)(x^2 + x + 1)$$

que no sea $\zeta = 1$, por lo que será, de hecho, raíz del polinomio $x^2 + x + 1$. Este polinomio es irreducible en $\mathbb{Q}(x)$, por lo que podemos aplicarle el teorema 4.1. No obstante, para no repetir una vez más unos argumentos que ya hemos empleado dos veces, vamos a probar un resultado general sobre la aplicación de dicho teorema a un anillo de polinomios:

Teorema 6.1 *Sea k un cuerpo y $p(x)$ un polinomio irreducible en $k[x]$. Sea K el anillo de clases de restos en $k[x]$ módulo $p(x)$. Entonces:*

1. *K es un cuerpo.*
2. *k se identifica con un subcuerpo de K identificando cada $a \in k$ con la clase de restos $\bar{a}$.*

3. *Si llamamos $\alpha = \bar{x}$, entonces $p(\alpha) = 0$.*

4. *Si $p(x)$ tiene grado n, todo elemento de K se expresa de forma única como*

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}, \qquad \text{con } a_0, \ldots, a_{n-1} \text{ en } k.$$

5. *Si K' es un cuerpo que contiene a k y α' es un elemento de K' tal que $p(\alpha') = 0$, entonces existe una única aplicación $f : K \longrightarrow K'$ que cumple:*

   (a) *f permite identificar a K con un subcuerpo de K', es decir, hace corresponder elementos distintos de K con elementos distintos de K' y además:*

$$f(u + v) = f(u) + f(v), \qquad f(uv) = f(u)f(v).$$

   (b) *$f(u) = u$ para todo u de k.*

   (c) *$f(\alpha) = \alpha'$.*

Demostración: Sabemos que K es un cuerpo por el teorema 4.1.

Un elemento arbitrario de K es de la forma $\overline{q(x)}$, para cierto polinomio $q(x)$ en $k[x]$. Podemos dividir $q(x) = p(x)c(x) + r(x)$, donde $r(x)$ tiene grado menor que n, pero entonces $\overline{q(x)} = \overline{r(x)}$. Si $r(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1}$, con los a_i en k, tenemos que

$$\overline{q(x)} = \bar{a}_0 + \bar{a}_1\alpha + \cdots + \bar{a}_{n-1}\alpha^{n-1}.$$

La expresión es única, pues si

$$\bar{a}_0 + \bar{a}_1\alpha + \cdots + \bar{a}_{n-1}\alpha^{n-1} = \bar{b}_0 + \bar{b}_1\alpha + \cdots + \bar{b}_{n-1}\alpha^{n-1},$$

llamando $r'(x) = b_0 + b_1x + \cdots + b_{n-1}x^{n-1}$, tenemos que $\overline{r(x)} = \overline{r'(x)}$, luego $r(x) - r'(x) = p(x)h(x)$, pero el polinomio de la izquierda tiene grado $< n$ y el de la derecha $\geq n$ salvo si $h(x) = 0$, luego tiene que ser $r(x) = r'(x)$, luego $a_i = b_i$ para todo i.

En particular, si a y b son elementos de k y se cumple que $\bar{a} = \bar{b}$, entonces $a = b$, lo cual demuestra 2). Escribiendo ahora a_i en lugar de $\bar{a}_i$, hemos demostrado 4) y 3) es inmediato, pues si $p(x) = c_nx^n + \cdots c_1x + c_0$, la igualdad $\overline{c(x)} = \bar{0}$ equivale a

$$\overline{c_nx^n + \cdots c_1x + c_0} = c_n\alpha^n + \cdots + c_1\alpha + c_0 = 0,$$

es decir, a que $p(\alpha) = 0$.

Para probar 4) consideramos $c : k[x] \longrightarrow K'$ dada por $c(q(x)) = q(\alpha')$, que claramente cumple

$$c(q_1(x) + q_2(x)) = c(q_1(x)) + c(q_2(x)), \qquad c(q_1(x)q_2(x)) = c(q_1(x))c(q_2(x)).$$

Si $\overline{q_1(x)} = \overline{q_2(x)}$, entonces $q_1(x) - q_2(x) = p(x)h(x)$, luego

$$c(q_1(x)) - c(q_2(x)) = c(p(x)h(x)) = p(\alpha')h(\alpha') = 0,$$

luego $c(q_1(x)) = c(q_2(x))$.

Esto significa que c da el mismo resultado sobre todos los elementos de una clase de restos módulo $p(x)$, por lo que podemos definir $f : K \longrightarrow K'$ mediante $f(\overline{q(x)}) = c(q(x))$, y claramente

$$f(u+v) = f(u) + f(v), \qquad f(uv) = f(u)f(v),$$

$f(\alpha) = f(\bar{x}) = c(x) = \alpha'$ y si a está en k, entonces $f(a) = c(a) = a$. Además, si $f(u) = f(v)$, entonces $f(u-v) = 0$, y esto implica que $u - v = 0$, pues en caso contrario tendríamos que

$$f(1) = f((u-v)(u-v)^{-1}) = f(u-v)f((u-v)^{-1}) = 0,$$

pero en realidad $f(1) = 1$. ∎

Los resultados básicos sobre los cuerpos $\mathbb{Q}(i)$ y $\mathbb{Q}\big(\sqrt{-2}\,\big)$ recogidos en los resúmenes 4.1 y 5.1 son casos particulares del teorema anterior. Ahora estamos interesados en aplicarlo al polinomio $p(x) = x^2 + x + 1$, y el resultado es el siguiente:

Definición 6.2 Llamaremos $\mathbb{Q}(\zeta)$ al cuerpo dado por el teorema anterior para $k = \mathbb{Q}$ y $p(x) = x^2 + x + 1$. Llamamos $\zeta = \bar{x}$, de modo que $\mathbb{Q}(\zeta)$ es un cuerpo cuyos elementos se expresan de forma única en la forma $a + b\zeta$, donde a y b son números racionales y $\zeta^2 = -\zeta - 1$. En particular $\zeta^3 = 1$.

Esto determina las operaciones en $\mathbb{Q}(\zeta)$:

$$(a + b\zeta) + (c + d\zeta) = (a + c) + (b + d)\zeta,$$

$$(a + b\zeta)(c + d\zeta) = ac + (ad + bc)\zeta + bd\zeta^2 = (ac - bd) + (ad + bc - bd)\zeta.$$

El teorema anterior prueba también que en $\mathbb{Q}(\zeta)$ podemos definir una conjugación, pero a la vez nos da una visión más profunda de lo que debemos entender por "conjugar". En los ejemplos que habíamos manejado hasta ahora el conjugado de i era $-i$ y el conjugado de $\sqrt{-2}$ era $-\sqrt{-2}$, lo cual puede llevarnos a generalizar ingenuamente que "conjugar" es cambiarle el signo a la "parte imaginaria" de un número, y en general no es así. La razón de fondo por la que $-i$ es el conjugado de i es porque i es raíz del polinomio $x^2 + 1$ y $-i$ es *la otra raíz* de dicho polinomio, y lo mismo sucede con $-\sqrt{-2}$, que es *la otra raíz* del polinomio $x^2 + 2$.

En cambio, *la otra raíz* de x^2+x+1 no es $-\zeta$, sino ζ^2. En efecto, si dividimos

$$\begin{array}{c|ccc} & 1 & 1 & 1 \\ \zeta & & \zeta & -1 \\ \hline & 1 & -\zeta^2 & 0 \end{array}$$

vemos que $x^2+x+1 = (x-\zeta)(x-\zeta^2)$. Por lo tanto, el teorema anterior afirma que podemos definir el conjugado de un elemento $z = a+b\zeta$ de $\mathbb{Q}(\zeta)$ como

$$\bar{z} = a + b\zeta^2 = a - b - b\zeta,$$

de modo que se cumplen las mismas propiedades básicas que tiene la conjugación en $\mathbb{Q}(i)$ o en $\mathbb{Q}\big(\sqrt{-2}\big)$.

Ejercicio: Comprobar que si definiéramos $\overline{a+b\zeta} = a-b\zeta$ no se cumpliría $\overline{z_1 z_2} = \bar{z}_1\bar{z}_2$.

Hay un par de propiedades que son triviales cuando "conjugar es cambiar el signo a la parte imaginaria", que en este contexto requieren una ligera comprobación. Por ejemplo:

$$\bar{\bar{z}} = \overline{a+b\zeta^2} = a + b\bar{\zeta}^2 = a + b\zeta^4 = a + b\zeta.$$

Y, por otra parte, es fácil ver también que $\bar{z} = z$ si y sólo si z es un número racional.

A su vez, podemos definir la norma $\mathrm{N} : \mathbb{Q}(\zeta) \longrightarrow \mathbb{Q}$ mediante $\mathrm{N}(z) = z\bar{z}$ o, explícitamente

$$\mathrm{N}(a+b\zeta) = (a+b\zeta)(a+b\zeta^2) = a^2+b^2+ab\zeta+ab\zeta^2 = a^2-ab+b^2.$$

La expresión $\mathrm{N}(z) = z\bar{z}$ implica trivialmente que $\mathrm{N}(z_1z_2) = \mathrm{N}(z_1)\,\mathrm{N}(z_2)$, así como que la norma sólo se anula en 0. Notemos además que

$$\mathrm{N}(a+b\zeta) = (a-b/2)^2 + 3b^2/4 \geq 0.$$

La norma nos permite calcular cocientes fácilmente:

$$\frac{3+2\zeta}{2+\zeta} = \frac{(3+2\zeta)(2+\zeta^2)}{2^2-2+1} = \frac{8+4\zeta+3\zeta^2}{3} = \frac{5+\zeta}{3} = \frac{5}{3}+\frac{1}{3}\zeta.$$

Definimos el anillo $\mathbb{Z}[\zeta]$ de los *enteros de Eisenstein* como el formado por los números de la forma $a+b\zeta$ con a y b en $\mathbb{Z}$. Las expresiones explícitas para la suma y el producto muestran que la suma y el producto de enteros de Eisenstein es un entero de Eisenstein, por lo que realmente tenemos un anillo, un dominio íntegro, de hecho, cuyo cuerpo de cocientes es $\mathbb{Q}(\zeta)$.

Nota Hemos visto que todo elemento de $\mathbb{Q}(\zeta)$ se expresa de forma única como $a+b\zeta$, donde a y b son números racionales. Sin embargo, a la hora de operar con estos números, es más práctico expresarlos en la forma $c\zeta^2+b\zeta+a$. Así la expresión no es única, pero es fácil reconocer cuándo dos expresiones corresponden al mismo número:

$$\textit{Se cumple } c\zeta^2+b\zeta+a = c'\zeta^2+b'\zeta+a' \textit{ si y sólo si } c-c' = b-b' = a-a'.$$

En efecto, si llamamos $q(x) = cx^2+bx+c$ y $q'(x) = c'x^2+b'x+c'$, tenemos que $\overline{q(x)} = \overline{q'(x)}$, lo que equivale a que exista un polinomio $h(x)$ tal que

$$q(x) - q'(x) = (x^2+x+1)h(x),$$

pero el miembro izquierdo tiene grado ≤ 2, luego $h(x)$ tiene que ser una constante h, luego la igualdad equivale a que exista un número racional h tal que

$$(c-c')x^2+(b-b')x+c-c'=h(x^2+x+1),$$

que a su vez equivale a que $c-c'=b-b'=a-a'$. ∎

Por ejemplo:

$$3\zeta^2+2\zeta-5=9z^2+8\zeta+1=-\zeta-8.$$

Notemos que para expresar un número $c\zeta^2+b\zeta+a$ en la forma reducida $b'\zeta+a'$ sólo tenemos que restar c a todos sus coeficientes.

La ventaja de permitir la presencia de potencias de ζ es que las sumas se calculan con la misma facilidad, mientras que para multiplicar basta tener en cuenta la relación $\zeta^3=1$, que es mucho mas cómoda de manejar que $\zeta^2=-\zeta-1$.

He aquí un ejemplo de suma y de multiplicación calculadas mediante estas expresiones no reducidas:

$$\begin{array}{r} 2\zeta^2-3\zeta+4 \\ +\ 5\zeta^2+7\zeta-7 \\ \hline 7\zeta^2+4\zeta-3 \end{array} \qquad \begin{array}{r} 2\zeta^2\ -3\zeta\ +4 \\ \times\ 5\zeta^2\ +7\zeta\ -7 \\ \hline -14\zeta^2-21\zeta-28 \\ -21\zeta^2+28\zeta+14 \\ 20\zeta^2+10\zeta-15 \\ \hline -15\zeta^2+17\zeta-29 \end{array}$$

Una vez hechas todas las operaciones, es fácil expresar el resultado en forma reducida: $-3\zeta-10$ para la suma y $32\zeta-14$ para el producto.

Un ligero inconveniente es que puede costar un poco más identificar a los enteros de Eisenstein, pues, por ejemplo

$$\frac{1}{2}\zeta^2+\frac{3}{2}\zeta-\frac{5}{2}$$

es entero. No obstante, cuando operamos con enteros de Eisenstein no tenemos necesidad de considerar nunca expresiones con coeficientes fraccionarios. ∎

Teorema 6.3 *El anillo $\mathbb{Z}[\zeta]$ es un dominio euclídeo tomando com norma euclídea la norma algebraica $\mathrm{N}:\mathbb{Z}[\zeta]\longrightarrow\mathbb{N}$.*

Demostración: Teniendo en cuenta que las normas de los enteros de Eisenstein son números naturales, es evidente que si α y β son enteros de Eisenstein no nulos, entonces $\mathrm{N}(\alpha)\leq\mathrm{N}(\alpha)\,\mathrm{N}(\beta)=\mathrm{N}(\alpha\beta)$.

Tomemos ahora dos enteros de Eisenstein Δ y δ con $\delta\neq 0$. Consideramos el cociente $\Delta/\delta=r+s\zeta$, donde r y s son números racionales. Llamamos x e y a los enteros racionales más cercanos a r y s, respectivamente, de modo que $|r-x|$, $|s-y|\leq 1/2$, con lo que, llamando $\gamma=x+y\zeta$, $\epsilon=\Delta-\delta\gamma$, tenemos que

$$\mathrm{N}(\epsilon/\delta)=\mathrm{N}(\Delta/\delta-\gamma)=(r-x)^2-(r-x)(s-y)+(s-y)^2\leq\frac{3}{4}<1,$$

luego $\mathrm{N}(\epsilon)<\mathrm{N}(\delta)$. ∎

Así pues, ahora sabemos que los enteros de Eisenstein poseen una aritmética análoga a la de los enteros racionales, la de los enteros de Gauss o la de los enteros $\mathbb{Z}\big[\sqrt{-2}\,\big]$, pero tenemos que estudiar sus particularidades.

Unidades Como en los otros ejemplos de anillos de enteros que hemos estudiado, es inmediato que las unidades de $\mathbb{Z}[\zeta]$ son los enteros de Eisenstein que cumplen $\mathrm{N}(\alpha) = 1$. Haciendo $\alpha = a + b\zeta$, esto equivale a

$$a^2 - ab + b^2 = (a - b/2)^2 + 3b^2/4 = 1,$$

o también a

$$(2a - b)^2 + 3b^2 = 4.$$

Es claro entonces que tiene que ser $b = -1, 0, 1$ y, al calcular los valores posibles de a para cada uno de estos valores de b obtenemos:

Las unidades de $\mathbb{Z}[\zeta]$ *son* $\pm 1, \pm\zeta, \pm\zeta^2$.

Así pues, en el anillo de los enteros de Eisenstein, los asociados vienen en grupos de seis. Más concretamente, es claro que los asociados de $a + b\zeta + c\zeta^2$ son los seis números que se obtienen permutando cíclicamente los coeficientes o cambiando los signos de todos ellos.

Primos Si π es primo en $\mathbb{Z}[\zeta]$, entonces $\pi \mid \pi\bar{\pi} = \mathrm{N}(\pi)$, por lo que, descomponiendo $\mathrm{N}(\pi)$ en primos racionales, concluimos que π divide a uno de ellos. Así pues, todo primo de Eisenstein aparece en la descomposición en factores primos de un primo racional.

Además, si $\pi \mid p$, tenemos que $\mathrm{N}(\pi) \mid \mathrm{N}(p) = p^2$, lo que prueba que cada primo cuadrático divide a un único primo racional p (el que divide a su norma) y que $\mathrm{N}(\pi)$ tiene que ser p o p^2. Además, cada primo racional p se descompone a lo sumo en dos factores primos, que pueden ser asociados o no, lo que nos lleva a las definiciones usuales:

Definición 6.4 Si p es un primo racional, se dice que:

1. p se *ramifica* en $\mathbb{Z}[\zeta]$ si se descompone como $p = \pm\pi^2$, donde π es primo de norma p.
2. p se *escinde* en $\mathbb{Z}[\zeta]$ si se descompone en dos factores primos no asociados $p = \pi_1\pi_2$ (de norma p).
3. p se *conserva* (primo) en $\mathbb{Z}[\zeta]$ si es primo en $\mathbb{Z}[\zeta]$ (y entonces tiene norma p^2).

No hay más posibilidades. Por ejemplo, si llamamos $\lambda = 1 - \zeta$, tenemos que $\mathrm{N}(\lambda) = 3$, por lo que

$$3 = (1 - \zeta)(1 - \zeta^2)$$

es una descomposición de 3 en factores primos, pero no es cierto que 3 se escinda, pues

$$3 = (1 - \zeta)(1 - \zeta^2) = (1 - \zeta)(1 - \zeta)(1 + \zeta) = -\zeta^2(1 - \zeta)^2,$$

luego en realidad 3 se ramifica.

Exactamente el mismo argumento empleado para los enteros de Gauss nos da el teorema siguiente:

Teorema 6.5 *Sea p un primo racional. Entonces:*

1. *Si $x^2+x+\bar{1} = (x-\bar{c})^2$ en $\mathbb{Z}_p[x]$, entonces $p = \epsilon\pi^2$, donde $\pi = (p, \zeta - c)$ es un primo de Eisenstein que cumple $\zeta \equiv c \pmod{\pi}$.*

2. *Si $x^2+x+\bar{1} = (x-\bar{c}_1)(x-\bar{c}_2)$ en $\mathbb{Z}_p[x]$, donde $c_1 \not\equiv c_2 \pmod{p}$, entonces $p = \pi_1\pi_2$, donde $\pi_i = (p, \zeta - c_i)$ son primos de Eisenstein no asociados tales que $\zeta \equiv c_j \pmod{\pi_j}$.*

3. *Si $x^2+x+\bar{1}$ es irreducible en $\mathbb{Z}_p[x]$, entonces p se conserva primo y ζ no es congruente módulo p con ningún entero racional.*

Así, como el polinomio x^2+x+1 tiene discriminante $\Delta = -3$, concluimos que 3 es el único primo racional que se ramifica en $\mathbb{Z}[\zeta]$. Como x^2+x+1 no tiene raíces en $\mathbb{Z}_2$, concluimos que 2 se conserva primo. Para los primos $p > 3$, tenemos que p se escinde o se conserva primo según si[1] -3 es un resto cuadrático o un resto no cuadrático módulo p. Según el teorema 3.39, podemos afirmar, más precisamente, que los primos que se escinden son los que cumplen $p \equiv 1 \pmod{3}$ y los que se conservan los que cumplen $p \equiv -1 \pmod{3}$.

Por ejemplo, es fácil ver que x^2+x+1 no tiene raíces en $\mathbb{Z}_5$, luego 5 también se conserva primo. En cambio, $x^2+x+1 \equiv (x-2)(x+3) \pmod{7}$, luego 7 se escinde y sus factores primos son

$$\pi_1 = (7, \zeta - 2), \qquad \pi_2 = (7, \zeta + 3).$$

Como $\mathrm{N}(\zeta-2) = \mathrm{N}(\zeta+3) = 7$, ambos números son primos, luego son π_1 y π_2. Explícitamente:

$$7 = (\zeta-2)(\zeta^2-2) = (\zeta-2)(-\zeta-3) = -(\zeta-2)(\zeta+3).$$

Notemos que nuevamente nos encontramos con una "reciprocidad cuadrática", en el sentido de que el problema de si el discriminante $\Delta = -3$ es un resto cuadrático módulo un primo impar p resulta ser equivalente a la congruencia $p \equiv 1 \pmod{3}$, pero si además nos damos cuenta de que 1 es el único resto cuadrático (no nulo) módulo 3, podríamos expresar el resultado en estos términos:

$$\left(\frac{-3}{p}\right) = \left(\frac{p}{3}\right).$$

Hemos probado esta igualdad a partir del teorema 3.37, pero sucede que esa demostración no es generalizable a otros casos similares. Por eso es interesante

[1] Notemos que este criterio no es válido para $p = 2$ porque la fórmula que da las raíces de un polinomio de segundo grado no es válida en cuerpo de característica 2. De hecho, $\Delta = -3$ es un resto cuadrático módulo 2, pero eso no significa que el polinomio x^2+x+1 tenga raíces en $\mathbb{Z}_2$.

ver otra demostración alternativa basada en la misma técnica que usamos para demostrar 5.11 en la sección 5.3. Allí nos basamos en la expresión

$$\sqrt{2} = \zeta + \zeta^{-1}$$

que relaciona una raíz cuadrada de 2 con una raíz octava primitiva de la unidad. Nosotros tenemos una relación entre una raíz cuadada de -3 y una raíz cúbica primitiva de la unidad, a saber:

$$-3 = \zeta^2(1-\zeta)^2.$$

Podemos expresarla de forma más simple si operamos un poco:

$$-3 = (\zeta - \zeta^2)^2.$$

con lo que $\sqrt{-3} = \zeta - \zeta^2$.

Lo que hicimos en la sección 5.11 fue comprobar que la relación valía en realidad para cualquier raíz primitiva de la unidad en cualquier cuerpo, en particular, en un cuerpo que contuviera a $\mathbb{Z}_p$. Aquí vamos a hacer lo mismo.

Aplicando el teorema 6.1 (o simplemente 4.1) a un factor irreducible del polinomio $x^2 + x + \bar{1}$ en $\mathbb{Z}_p$ (donde $p > 3$ es primo, y esta condición hace que 1 no sea raíz del polinomio), obtenemos un cuerpo k que contiene a $\mathbb{Z}_p$ en el cual existe un ζ que cumple $\zeta^2 + \zeta + \bar{1} = 0$. En particular $\zeta \neq \bar{1}$, pero

$$\zeta^3 = \zeta\zeta^2 = \zeta(-\zeta - \bar{1}) = -\zeta^2 - \zeta = \bar{1}.$$

Y además

$$(\zeta - \zeta^2)^2 = \zeta^2 + \zeta - 2 = -3.$$

Por lo tanto, podemos llamar $\sqrt{-3} = \zeta - \zeta^2$, que ahora es un elemento de k.

Tenemos que $(3/p) = 1$ si y sólo si el polinomio $x^2 + \bar{3}$ tiene sus raíces en $\mathbb{Z}_p$, pero, como $\sqrt{-3}$ es una de dichas raíces, esto equivale a que $\sqrt{-3}$ esté en $\mathbb{Z}_p$. Pero, tal y como razonamos en la sección 5.11, los elementos de $\mathbb{Z}_p$ son precisamente las raíces del polinomio $x^p - x$, luego $(-3/p) = 1$ equivale a que $(\sqrt{-3})^p = \sqrt{-3}$. Ahora bien:

$$(\sqrt{-3})^p = (\zeta - \zeta^2)^p = \zeta^p - \zeta^{2p} = \begin{cases} \zeta - \zeta^2 & \text{si } p \equiv \ \ 1 \text{ (mód 3)}, \\ \zeta^2 - \zeta & \text{si } p \equiv -1 \text{ (mód 3)} \end{cases} = \left(\frac{p}{3}\right)\sqrt{-3}.$$

Por lo tanto $(-3/p) = 1$ si y sólo si $(p/3) = 1$. ∎

Ejercicio: Caracterizar los números naturales que pueden expresarse en la forma $x^2 - xy + y^2$. Dar una fórmula explícita que muestre que el producto de dos números de esta forma es también de esta forma.

Ejemplo *La ecuación $y^2 + 3 = x^3 - x$ no tiene soluciones enteras.*

Observamos que $x^3 - x$ es siempre par, por lo que y tiene que ser impar. Entonces $y^2 + 3 \equiv 4$ (mód 8). Si x es impar, entonces $8 \mid x^2 - 1$ y tenemos una

contradicción, luego x es par y x^2-1 es impar. Como $4 \mid y^2+3 = x(x^2-1)$, concluimos que $4 \mid x$. Más aún, $8 \nmid x$. Escribimos la ecuación en la forma

$$y^2+3 = 4(x-1)(x+1)(x/4),$$

donde los tres últimos factores son impares. Como $x \equiv x/4 \pmod 3$, uno de los tres últimos factores es $\equiv -1 \pmod 3$ (esto lo cumple necesariamente uno de cada tres números consecutivos $x-1$, x, $x+1$). Dicho factor debe tener un factor primo $p \equiv -1 \pmod 3$, pero entonces $p \mid y^2+3$, lo que significa que $(-3/p) = 1$, con lo que tenemos una contradicción. ∎

El ejercicio siguiente permitirá al lector concebir la demostración precedente desde una perspectiva más amplia:

Ejercicio: Sea $p \neq 5$ un primo impar. Construir un cuerpo que contenga a $\mathbb{Z}_p$ que contenga una raíz quinta primitiva de la unidad ω, es decir, una raíz del polinomio $x^4+x^3+x^2+x+1$. Comprobar que $\omega^5 = 1$ y $\omega \neq 1$. Sea

$$\sqrt{5} = \left(\frac{1}{5}\right)\omega + \left(\frac{2}{5}\right)\omega^2 + \left(\frac{3}{5}\right)\omega^3 + \left(\frac{4}{5}\right)\omega^4 = \omega - \omega^2 - \omega^5 + \omega^4.$$

Comprobar que, como sugiere el nombre, $(\sqrt{5})^2 = 5$. Probar que

$$(\sqrt{5})^p = \left(\frac{p}{5}\right)\sqrt{5}.$$

deducir que $(5/p) = (p/5)$.

6.2 El anillo $\mathbb{Z}[\sqrt{-3}]$

Por definición —o por construcción— ζ es una raíz del polinomio x^2+x+1, pero podemos resolver la ecuación $x^2+x+1 = 0$ mediante la fórmula para ecuaciones de segundo grado, que nos da que sus raíces son

$$\frac{-1 \pm \sqrt{-3}}{2}.$$

Por otro lado tenemos que dichas raíces son ζ y ζ^2, luego podemos convenir que

$$\zeta = \frac{-1+\sqrt{-3}}{2}, \qquad \zeta^2 = \frac{-1-\sqrt{-3}}{2}.$$

Esto no es algo que tenga que ser demostrado, sino que al establecer estas igualdades estamos decidiendo a cuál de las dos raíces cuadradas de -3 en $\mathbb{Z}[\zeta]$ llamamos $\sqrt{-3}$ y a cuál llamamos $-\sqrt{-3}$. Recíprocamente:

$$\sqrt{-3} = 1+2\zeta, \qquad -\sqrt{-3} = -1-2\zeta = 1+2\zeta^2.$$

Estas expresiones muestran que $-\sqrt{-3}$ es el conjugado de $\sqrt{-3}$.

Ahora, todo elemento de $\mathbb{Q}(\zeta)$ es de la forma

$$a + b\zeta = a + b\frac{-1+\sqrt{-3}}{2} = \frac{2a-b}{2} + \frac{b}{2}\sqrt{-3} = c + d\sqrt{-3},$$

donde a y b (luego también c y d) son números racionales. La relación es

$$\frac{2a-b}{2} = c, \qquad \frac{b}{2} = d,$$

pero esto puede invertirse:

$$a = c + d \qquad b = 2d,$$

de donde podemos concluir que todo elemento de $\mathbb{Q}(\zeta)$ se expresa de forma única como $c+d\sqrt{-3}$, donde c y d son números racionales arbitrarios. En estos términos,

$$\overline{c + d\sqrt{-3}} = c - d\sqrt{-3},$$

luego

$$\mathrm{N}(c + d\sqrt{-3}) = (c + d\sqrt{-3})\overline{(c + d\sqrt{-3})} = c^2 + 3d^2.$$

A su vez, esto se interpreta como que, si construimos un cuerpo $\mathbb{Q}\big(\sqrt{-3}\,\big)$ aplicando el teorema 6.1 al polinomio $x^2 + 3$, el cuerpo que obtenemos no es sino $\mathbb{Q}(\zeta)$.

Más precisamente, tenemos que $\mathbb{Q}(\zeta)$ contiene una raíz $\sqrt{-3}$ del polinomio $x^2 + 3$, luego el teorema 6.1 nos da una aplicación $f : \mathbb{Q}\big(\sqrt{-3}\,\big) \longrightarrow \mathbb{Q}(\zeta)$ que nos permite identificar ambos cuerpos. Por lo tanto, podemos hablar del cuerpo $\mathbb{Q}\big(\sqrt{-3}\,\big) = \mathbb{Q}(\zeta)$ sin que importe si lo consideramos construido a partir de un polinomio o de otro.

Ahora bien, si —análogamente a como hemos hecho con $\mathbb{Z}[i]$ o $\mathbb{Z}\big[\sqrt{-2}\,\big]$— consideramos el anillo $\mathbb{Z}\big[\sqrt{-3}\,\big]$ de enteros cuadráticos formado por los números de la forma $a+b\sqrt{-3}$, con a y b enteros racionales, ya no es cierto que se cumpla $\mathbb{Z}\big[\sqrt{-3}\,\big] = \mathbb{Z}[\zeta]$. Concretamente, tenemos las inclusiones

$$\mathbb{Z}\big[\sqrt{-3}\,\big] \subsetneq \mathbb{Z}[\zeta] \subsetneq \mathbb{Q}(\zeta) = \mathbb{Q}\big(\sqrt{-3}\,\big).$$

Por ejemplo, tenemos que ζ no está en $\mathbb{Z}\big[\sqrt{-3}\,\big]$, pues su expresión

$$\zeta = -\frac{1}{2} + \frac{1}{2}\sqrt{-3}$$

no tiene coeficientes enteros. Más precisamente, un elemento arbitrario de $\mathbb{Z}[\zeta]$ es

$$a + b\zeta = a + b\frac{-1+\sqrt{-3}}{2} = a - \frac{b}{2} + \frac{b}{2}\sqrt{-3}$$

y así vemos que está en $\mathbb{Z}\big[\sqrt{-3}\,\big]$ si y sólo si $2 \mid b$.

Notemos que la norma de $\mathbb{Q}(\zeta)$ se restringe a $\mathrm{N} : \mathbb{Z}\big[\sqrt{-3}\,\big] \longrightarrow \mathbb{N}$ dada por

$$\mathrm{N}(c + d\sqrt{-3}) = c^2 + 3d^2.$$

Por ejemplo, considerando esta norma es inmediato comprobar que las unidades de $\mathbb{Z}\big[\sqrt{-3}\,\big]$ son simplemente ± 1, igual que las de $\mathbb{Z}\big[\sqrt{-2}\,\big]$.

Sin embargo, si ahora el lector trata de adaptar el argumento que hemos empleado para demostrar que $\mathbb{Z}[i]$, $\mathbb{Z}\big[\sqrt{-2}\,\big]$ o $\mathbb{Z}[\zeta]$ son dominios euclídeos, verá que no lo consigue, porque sucede que $\mathbb{Z}\big[\sqrt{-3}\,\big]$ no es un dominio euclídeo. Más aún: ¡no es un dominio de factorización única! En efecto:

$$2 \cdot 2 = 4 = (1 + \sqrt{-3})(1 - \sqrt{-3}).$$

Observemos que ninguno de los cuatro factores es una unidad de $\mathbb{Z}\big[\sqrt{-3}\,\big]$, pues éstas son ± 1, y todos ellos son irreducibles, pues tienen norma 4 y, si pudieran descomponerse en factores irreducibles, tendrían que ser dos factores de norma 2, pero es inmediato que en $\mathbb{Z}\big[\sqrt{-3}\,\big]$ no hay elementos de norma 2. Así pues, tenemos dos descomposiciones distintas de 4 en factores irreducibles, lo que niega la factorización única.

Notemos en particular que los cuatro factores son irreducibles, pero no son primos. Por ejemplo, 2 divide al producto de la derecha, pero no divide a ninguno de sus factores, y viceversa.

Observemos que esto no contradice la factorización única de $\mathbb{Z}[\zeta]$, pues en este anillo tenemos que

$$1 + \sqrt{-3} = -\frac{-1 - \sqrt{-3}}{2} \cdot 2 = -\zeta^2 \cdot 2, \quad 1 - \sqrt{-3} = -\frac{-1 + \sqrt{-3}}{2} \cdot 2 = -\zeta \cdot 2,$$

luego las dos descomposiciones son "la misma", ya que cada factor de una es asociado a un factor de la otra.

Nota Conviene señalar que podíamos haber asegurado que $\mathbb{Z}\big[\sqrt{-3}\,\big]$ no tiene factorización única sin haber hecho cálculo alguno. Esto es consecuencia del teorema 2.28. El polinomio $x^2 + x + 1$ tiene sus coeficientes en $\mathbb{Z}\big[\sqrt{-3}\,\big]$ y tiene una raíz ζ en $\mathbb{Q}\big(\sqrt{-3}\,\big)$, luego si $\mathbb{Z}\big[\sqrt{-3}\,\big]$ fuera un dominio de factorización única podríamos concluir que $\zeta \in \mathbb{Z}\big[\sqrt{-3}\,\big]$, lo cual es falso. ∎

Por otra parte, la violación que hemos encontrado de la factorización única en $\mathbb{Z}\big[\sqrt{-3}\,\big]$ es esencialmente la única que se da. Para probarlo observamos algunos hechos:

1. *Todo elemento de $\mathbb{Z}[\zeta]$ tiene un asociado en $\mathbb{Z}\big[\sqrt{-3}\,\big]$.*

   En efecto, si $\alpha = a + b\zeta$ y b es par, entonces α ya está en $\mathbb{Z}\big[\sqrt{-3}\,\big]$. Si a es par entonces $\zeta^2\alpha = b + a\zeta^2 = b - a - a\zeta$ está en $\mathbb{Z}\big[\sqrt{-3}\,\big]$. Por último, si a y b son impares, entonces $\zeta\alpha = a\zeta + b\zeta^2 = -b + (a - b)\zeta$ está en $\mathbb{Z}\big[\sqrt{-3}\,\big]$.

2. *Si α, β son elementos no nulos de $\mathbb{Z}\big[\sqrt{-3}\big]$ y $\mathrm{N}(\alpha)$ es impar, entonces $\alpha \mid \beta$ en $\mathbb{Z}\big[\sqrt{-3}\big]$ si y sólo si $\alpha \mid \beta$ en $\mathbb{Z}[\zeta]$.*

   En efecto, si $\alpha = a+b\zeta$ y $\beta = c+d\zeta$, con b y d, pares, entonces a es impar, o de lo contrario $2 \mid \alpha$ (en $\mathbb{Z}[\zeta]$) y $\mathrm{N}(\alpha)$ sería par. Supongamos que

   $$c + d\zeta = (a + b\zeta)(u + v\zeta) = (au - bv) + (av + bu - bv)\zeta.$$

   Entonces $2 \mid av$, luego $2 \mid v$, y así $u + v\zeta$ está en $\mathbb{Z}\big[\sqrt{-3}\big]$.

3. *Un elemento π de $\mathbb{Z}\big[\sqrt{-3}\big]$ de norma impar es primo en $\mathbb{Z}\big[\sqrt{-3}\big]$ si y sólo si lo es en $\mathbb{Z}[\zeta]$.*

   Si π divide a un producto de elementos de $\mathbb{Z}\big[\sqrt{-3}\big]$, entonces divide a uno de los factores en $\mathbb{Z}[\zeta]$, luego también en $\mathbb{Z}\big[\sqrt{-3}\big]$.

4. *Todo elemento de $\mathbb{Z}\big[\sqrt{-3}\big]$ de norma impar se descompone en producto de factores primos en $\mathbb{Z}\big[\sqrt{-3}\big]$ de forma única salvo orden o asociación.*

   Si α es un elemento de $\mathbb{Z}\big[\sqrt{-3}\big]$ de norma impar, podemos descomponerlo en factores primos en $\mathbb{Z}[\zeta]$ y, eligiendo asociados adecuados, podemos tomarlos en $\mathbb{Z}\big[\sqrt{-3}\big]$. En principio esto nos da una descomposición de la forma

   $$\alpha = \epsilon\pi_1^{e_1} \cdots \pi_r^{e_r},$$

   donde ϵ es una unidad de $\mathbb{Z}[\zeta]$, pero $\alpha' = \pi_1^{e_1} \cdots \pi_r^{e_r}$ es un elemento de $\mathbb{Z}\big[\sqrt{-3}\big]$ de norma impar que divide a α en $\mathbb{Z}[\zeta]$, luego tiene que dividirlo en $\mathbb{Z}\big[\sqrt{-3}\big]$, luego ϵ tiene que estar en $\mathbb{Z}\big[\sqrt{-3}\big]$, luego $\epsilon = \pm 1$.

   La prueba del teorema fundamental de la aritmética muestra que dos descomposiciones en factores primos (no meramente irreducibles) de un mismo elemento son siempre iguales salvo orden y asociación.

5. *Si α es un elemento de $\mathbb{Z}\big[\sqrt{-3}\big]$ tal que $\mathrm{N}(\alpha) = 2^r m$, con m impar, entonces $\alpha = \beta\gamma$, donde $\mathrm{N}(\beta) = 2^r$ y $\mathrm{N}(\gamma) = m$. Además γ se descompone de forma única en factores primos, mientras que β se descompone como producto de factores ± 2, $\pm(1 + \sqrt{-3})$ y $\pm(1 - \sqrt{-3})$.*

   Descomponemos α es factores primos en $\mathbb{Z}[\zeta]$, eligiéndolos en $\mathbb{Z}\big[\sqrt{-3}\big]$, y llamamos γ al producto de todos los factores primos de norma impar. Es claro entonces que $\mathrm{N}(\gamma) = m$ y γ divide a α en $\mathbb{Z}[\zeta]$, luego también en $\mathbb{Z}\big[\sqrt{-3}\big]$. Esto significa que $\alpha = \beta\gamma$, para un cierto β en $\mathbb{Z}\big[\sqrt{-3}\big]$ tal que $\mathrm{N}(\beta) = 2^r$. Además, descomponiendo β en $\mathbb{Z}[\zeta]$ tenemos que $\beta = \epsilon 2^{r/2}$, donde ϵ es una unidad (y r es necesariamente par, porque no hay elementos de norma 2). Podemos suponer que $r \geq 1$. Entonces, si $\epsilon = \pm\zeta$, cambiamos $\epsilon 2$ por $\pm(-1 + \sqrt{-3})$ y si $\epsilon = \pm\zeta^2$ cambiamos $\epsilon 2$ por $\pm(-1 - \sqrt{-3})$.

Vemos así que, en principio, sólo es necesario que aparezca un factor irreducible de la forma $\pm(1 \pm \sqrt{-3})$, pero la relación

$$2 \cdot 2 = 4 = (1 + \sqrt{-3})(1 - \sqrt{-3}).$$

permite sustituir cada par de doses por dos factores de este tipo y la descomposición no es única a causa (únicamente) de esto.

El primer punto del razonamiento anterior tiene como consecuencia que un número natural es la norma de un elemento de $\mathbb{Z}\big[\sqrt{-3}\,\big]$ si y sólo si es la norma de un elemento de $\mathbb{Z}[\zeta]$. Por lo tanto:

Ejercicio: Caracterizar los números naturales de la forma $x^2 + 3y^2$.

6.3 El Último Teorema de Fermat para $p = 3$

En esta sección usaremos los enteros de Eisenstein para demostrar el Último Teorema de Fermat para el exponente $p = 3$, es decir, vamos a probar que la ecuación $x^3 + y^3 + z^3$ no tiene soluciones enteras no triviales (es decir, tales que $xyz \neq 0$).

Recordemos que $\lambda = \zeta - 1$ es primo y divide a 3, cuya factorización es $3 = -\zeta^2\lambda^2$.

Más en general, si restamos dos de las unidades 1, ζ, ζ^2 obtenemos un asociado de λ:

$$\zeta - 1 = \lambda, \quad \zeta^2 - 1 = -\zeta^2\lambda, \quad \zeta^2 - \zeta = \zeta\lambda.$$

En cambio, si las sumamos, obtenemos una unidad:

$$1 + \zeta = -\zeta^2, \quad 1 + \zeta^2 = -\zeta, \quad \zeta + \zeta^2 = -1.$$

Si sumamos una consigo misma obtenemos un asociado del primo 2:

$$1 + 1 = 2, \quad \zeta + \zeta = 2\zeta, \quad \zeta^2 + \zeta^2 = 2\zeta^2.$$

De aquí concluimos la primera de las propiedades siguientes:

- *Si dos unidades cumplen* $\epsilon_1 \equiv \epsilon_2 \pmod 3$, *tiene que ser* $\epsilon_1 = \epsilon_2$, *puesto que la diferencia sólo puede ser divisible entre* 2 *o entre* λ *salvo que sean la misma.*

- *Si* $\alpha = a + b\zeta$, *se cumple que* $\lambda \mid \alpha$ *si y sólo si* $3 \mid a + b$.

  En efecto, obviamente $\zeta \equiv 1 \pmod \lambda$, luego $a + b\zeta \equiv a + b \pmod \lambda$. Por lo tanto, si $\lambda \mid \alpha$, tenemos que $\lambda \mid a + b$, luego $3 = \mathrm{N}(\lambda) \mid (a + b)^2$, luego $3 \mid a + b$. Recíprocamente, si $3 \mid a + b$, entonces $\lambda \mid a + b$, luego $\lambda \mid \alpha$.

- *Si* $\lambda \nmid \alpha$, *entonces existe una unidad* ϵ *tal que* $\alpha \equiv \epsilon \pmod 3$.

  Como $3 \nmid a + b$, tiene que darse uno de los casos siguientes, con $e = \pm 1$:

  $$(a, b) = (3h + e, 3k), \ (3h, 3k + e), \ (3h + e, 3k + e),$$

  luego $\alpha \equiv e \pmod 3$, $\alpha \equiv e\zeta \pmod 3$, $\alpha \equiv e(1 + \zeta) \pmod 3$, respectivamente, y en cualquier caso $\alpha \equiv \epsilon \pmod 3$, para una cierta unidad ϵ.

- *Si $\lambda \nmid \alpha$, entonces $\alpha^3 \equiv \pm 1$ (mód 9).*

  Por la propiedad precedente sabemos que $\alpha = 3\beta + \epsilon$, luego

$$\alpha^3 = (3\beta + \epsilon)^3 \equiv \epsilon^3 = \pm 1 \text{ (mód 9)}.$$

Ahora mostramos la conexión entre la ecuación de Fermat y los enteros de Eisenstein. Observemos que -1, $-\zeta$ y $-\zeta^2$ son raíces distintas en el cuerpo $\mathbb{Q}(\zeta)$ del polinomio $t^3 + 1$, luego son todas sus raíces, y podemos factorizar:

$$t^3 + 1 = (t + 1)(t + \zeta)(t + \zeta^2).$$

En particular, si β, $\gamma \in \mathbb{Z}[\zeta]$ y $\gamma \neq 0$, sustituyendo $t = \beta/\gamma$ y multiplicando por γ^3 queda

$$\beta^3 + \gamma^3 = (\beta + \gamma)(\beta + \gamma\zeta)(\beta + \gamma\zeta^2),$$

que vale igualmente si $\gamma = 0$. Si multiplicamos el segundo factor por ζ y el tercero por ζ^2 en total hemos multiplicado por $\zeta^3 = 1$, luego la igualdad sigue siendo válida, pero pasa a tener la forma

$$\beta^3 + \gamma^3 = (\beta + \gamma)(\beta\zeta + \gamma\zeta^2)(\beta\zeta^2 + \gamma\zeta). \tag{6.1}$$

Esta segunda factorización tiene la propiedad adicional de que

$$\beta + \gamma + \beta\zeta + \gamma\zeta^2 + \beta\zeta^2 + \gamma\zeta = 0.$$

Con esto ya podemos probar:

Teorema 6.6 *La ecuación $x^3 + y^3 + z^3 = 0$ no tiene soluciones no triviales (es decir, con todas las variables no nulas) en $\mathbb{Z}[\zeta]$, en particular en $\mathbb{Z}$.*

Demostración: Supongamos que la ecuación tiene una solución no trivial $\alpha^3 + \beta^3 + \gamma^3 = 0$. Entonces existe una solución *primitiva,* es decir, una solución no trivial tal que α, β y γ sean primos entre sí dos a dos, pues si un primo ρ divide a dos de ellos, la ecuación hace que divida al tercero, luego α/ρ, β/ρ, γ/ρ son también una solución no trivial de la ecuación, y repitiendo este proceso un número finito de veces tenemos que llegar a una solución primitiva.

Observemos ahora que λ tiene que dividir a uno de los números α, β, γ (y sólo a uno, porque la solución es primitiva). En efecto, si λ no divide a ninguno de ellos, hemos visto que

$$\alpha^3 \equiv e_1 \text{ (mód 9)}, \quad \beta^3 \equiv e_2 \text{ (mód 9)}, \quad \gamma^3 \equiv e_3 \text{ (mód 9)},$$

donde $e_i = \pm 1$. Pero entonces[2] $e_1 + e_2 + e_3 \equiv 0$ (mód 9), lo cual es imposible.

Llamaremos *coordenada especial* de una solución primitiva (α, β, γ) a la que es divisible entre λ. Si la ecuación tiene solución, debería haber una solución primitiva en la que el exponente de λ en la coordenada especial fuera el menor número natural posible (no nulo). Sin embargo, vamos a probar que a partir de cualquier solución primitiva se puede encontrar otra tal que el exponente de λ en la coordenada especial sea menor, con lo que tendremos una contradicción.

[2]En principio, esta congruencia es en $\mathbb{Z}[\zeta]$, pero sabemos que m, n, $r \in \mathbb{Z}$ cumplen $m \equiv n$ (mód r) en $\mathbb{Z}[\zeta]$ si y sólo si lo cumplen en $\mathbb{Z}$.

Dada la simetría, no perdemos generalidad si suponemos que $\lambda \mid \alpha$. Entonces $\lambda^3 \mid \alpha^3$ y la ecuación nos da que $\lambda^3 \mid \beta^3 + \gamma^3$. Como λ no divide a β ni a γ, sabemos que $\beta^3 \equiv e \pmod 9$, $\gamma^3 \equiv f \pmod 9$, donde $e = \pm 1$, $f = \pm 1$.

Como $\lambda^2 \mid 3$, en particular $\lambda^3 \mid 9$, luego las congruencias anteriores valen también módulo λ^3, y por consiguiente

$$e + f \equiv \beta^3 + \gamma^3 \equiv 0 \pmod{\lambda^3}.$$

En particular, $3 \mid e + f$ y esto sólo es posible si $e + f = 0$ (pues la suma sólo puede ser 0, 2 o -2). Por lo tanto,

$$-\alpha^3 = \beta^3 + \gamma^3 \equiv e + f = 0 \pmod 9.$$

Así pues, $\lambda^4 \mid \alpha^3$, por lo que $\lambda^2 \mid \alpha$.

En definitiva, con esto hemos demostrado que la coordenada especial debe ser divisible entre λ, no una, sino al menos dos veces.

Ahora consideramos la factorización de $-\alpha^3 = \beta^3 + \gamma^3$ dada por (6.1). Llamemos

$$\eta_1 = \beta + \gamma, \quad \eta_2 = \beta\zeta + \gamma\zeta^2, \quad \eta_3 = \beta\zeta^2 + \gamma\zeta,$$

de modo que $-\alpha^3 = \eta_1\eta_2\eta_3$ y $\eta_1 + \eta_2 + \eta_3 = 0$.

Observamos que $\eta_1 \equiv \eta_2 \equiv \eta_3 \pmod\lambda$, pues los tres son congruentes con $\beta + \gamma$. Como $\lambda^3 \mid \eta_1\eta_2\eta_3$, en principio λ divide a uno de los tres, pero como son congruentes, $\lambda \mid \eta_i$ para los tres valores $i = 1, 2, 3$.

Sea $\eta_i' = \eta_i/\lambda$ y vamos a probar que η_1', η_2' y η_3' son primos entre sí dos a dos. En efecto, si un primo ρ divide a dos de ellos, por ejemplo a η_1' y η_2', entonces $\lambda\rho \mid \eta_1$ y $\lambda\rho \mid \eta_2$, luego

$$\lambda\rho \mid \eta_2 - \zeta^2\eta_1 = \beta(\zeta - \zeta^2) = -\zeta\beta\lambda, \qquad \lambda\rho \mid \eta_2 - \zeta\eta_1 = \gamma(\zeta^2 - \zeta) = \zeta\gamma\lambda,$$

luego $\rho \mid \gamma$ y $\rho \mid \beta$, cuando β y γ son primos entre sí. Igualmente llegamos a una contradicción si suponemos que η_1' y η_3' o η_2' y η_3' no son primos entre sí.

Tenemos, pues, que $-\alpha^3 = \lambda^3\eta_1'\eta_2'\eta_3'$ y $\eta_1' + \eta_2' + \eta_3' = 0$. Sea $\eta = \alpha/\lambda$, que sigue cumpliendo $\lambda \mid \eta$, pues hemos visto que λ divide al menos dos veces a α. Entonces $-\eta^3 = \eta_1'\eta_2'\eta_3'$ y ahora razonamos como sigue: cada primo aparece con exponente múltiplo de 3 en el miembro izquierdo de esta ecuación, luego también en el miembro derecho, pero como los tres factores no tienen primos en común, podemos afirmar que cada primo aparece con exponente múltiplo de 3 en cada uno de ellos, es decir, que, para $i = 1, 2, 3$, se cumple que $\eta_i' = \epsilon_i\theta_i^3$, para ciertos $\theta_i \in \mathbb{Z}[\zeta]$ y ciertas unidades ϵ_i. Así pues,

$$-\eta^3 = \epsilon_1\epsilon_2\epsilon_3\,\theta_1^3\,\theta_2^3\,\theta_3^3, \quad \epsilon_1\theta_1^3 + \epsilon_2\theta_2^3 + \epsilon_3\theta_3^3 = 0,$$

y los θ_i son primos entre sí dos a dos porque lo eran los η_i'.

Más aún, $\eta = \epsilon\theta_1\theta_2\theta_3$, para cierta unidad ϵ tal que $\epsilon^3 = -\epsilon_1\epsilon_2\epsilon_3$. Pero toda unidad elevada al cubo da ± 1, luego $e = \epsilon_1\epsilon_2\epsilon_3 = \pm 1$. Como $\lambda \mid \eta$, tenemos que λ divide a un θ_i, y sólo a uno, porque son primos entre sí dos a dos. Recapitulemos lo que tenemos hasta aquí:

Hemos encontrado una ecuación

$$\epsilon_1\theta_1^3 + \epsilon_2\theta_2^3 + \epsilon_3\theta_3^3 = 0 \tag{6.2}$$

en la que los θ_i son primos entre sí dos a dos y uno de ellos es divisible entre λ, los ϵ_i son unidades y $e = \epsilon_1\epsilon_2\epsilon_3 = \pm 1$, y más aún, si $\lambda \mid \theta_i$, entonces su exponente cumple

$$v_\lambda(\theta_i) = \frac{1}{3}v_\lambda(\eta_i') = \frac{1}{3}(v_\lambda(\eta_i) - 1) < \frac{1}{3}(v_\lambda(\eta_1) + v_\lambda(\eta_2) + v_\lambda(\eta_3)) = v_\lambda(\alpha).$$

Sólo falta ver que podemos eliminar las unidades de (6.2) para obtener una nueva solución primitiva en la que el exponente de λ en la coordenada especial sea menor que el de partida. En lo que queda de la prueba sólo nos vamos a apoyar en estos hechos, luego por simetría podemos suponer que $\lambda \mid \theta_3$.

Entonces $\theta_1^3 \equiv e$ (mód 9) y $\theta_2^3 \equiv f$ (mód 9), para ciertos $e = \pm 1$, $f = \pm 1$. Como $\lambda \mid \theta_3$, en particular $3 \mid \theta_3^3$, luego (6.2) nos da que

$$e\epsilon_1 + f\epsilon_2 \equiv \epsilon_1\theta_1^3 + \epsilon_2\theta_2^3 \equiv -\epsilon_3\theta_3^3 \equiv 0 \text{ (mód 3)}.$$

Según hemos visto en las observaciones previas a este teorema, esto implica que $e\epsilon_1 + f\epsilon_2 = 0$, luego $\epsilon_1 = \pm\epsilon_2$, luego $e = \pm\epsilon_1^2\epsilon_3$. Por consiguiente, si multiplicamos (6.2) por $\pm\epsilon_1^2$ obtenemos

$$\pm\epsilon_1^3\theta_1^3 + \epsilon_1^3\theta_2^3 + e\theta_3^3 = 0$$

y, teniendo en cuenta que $e = \pm 1 = e^3$, esto equivale a

$$(\pm\epsilon_1\theta_1)^3 + (\epsilon_1\theta_2)^3 + (e\theta_3)^3 = 0,$$

luego llamando $\alpha' = \pm\epsilon_1\theta_1$, $\beta' = \epsilon_1\theta_2$, $\gamma' = e\theta_3$, vemos que se trata de una solución primitiva de la ecuación de Fermat con $0 < v_\lambda(\gamma') < v_\lambda(\alpha)$ y tenemos la contradicción que buscábamos. ∎

6.4 El test de Pépin

En 1877 el matemático francés Édouard Lucas demostró que el sexto número de Fermat:

$$F_6 = 2^{64} + 1 = 18\,446\,744\,073\,709\,551\,617$$

no es primo. Esto puede probarse usando un resultado publicado en 1877 por el también matemático francés Jean François Théophile Pépin:

Teorema 6.7 (Test de Pépin) *El número de Fermat $F_n = 2^{2^n} + 1$ es primo si y sólo si $3^{(F_n-1)/2} \equiv -1$ (mód F_n).*

Demostración: Si F_n es primo, como

$$2^{2^n} + 1 \equiv (-1)^{2^n} + 1 = 2 \equiv -1 \text{ (mód 3)}, \quad 2^{2^n} + 1 \equiv 1 \text{ (mód 4)},$$

tenemos que

$$\left(\frac{3}{F_n}\right) = \left(\frac{F_n}{3}\right) = \left(\frac{-1}{3}\right) = -1,$$

y basta aplicar el criterio de Euler 5.8:

$$3^{(F_n-1)/2} \equiv -1 \text{ (mód } F_n).$$

Recíprocamente, si se cumple esta congruencia y F_n no es primo, tendrá un divisor primo tal que $q^2 \leq F_n$. Vamos a probar que se da la desigualdad contraria y así tendremos una contradicción.

Tenemos que $3^{(F_n-1)/2} \equiv -1$ (mód q), luego, por el teorema de Fermat, se cumple que $o_q(3) = F_n - 1$ y $F_n - 1 \mid q$, luego $F_n < q < q^2$. ∎

Así, para probar que F_6 no es primo, tendríamos que calcular los restos módulo F_6 de las potencias sucesivas

$$3, \quad 3^2, \quad 3^{2^2}, \quad 3^{2^3}, \quad \ldots \quad 3^{2^{63}},$$

cada una de las cuales es el cuadrado de la anterior. Lucas simplificó sustancialmente los cálculos trabajando en base 2. Para entender el procedimiento vamos a considerar el caso mucho más simple de estudiar la primalidad de $F_3 = 257$. Los cálculos necesarios se pueden representar así:

14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	3^{2^n}	n
													1	1	3	0
													1	1		
											+	1	1			
											1	0	0	1	9	1
											1	0	0	1		
							+	1	0	0	1					
								1	0	1	0	0	0	1	81	2
								1	0	1	0	0	0	1		
				1	0	1	0	0	0	1						
	+	1	0	1	0	0	0	1								
		1	1	0	0	1	1	0	1	0	0	0	0	1	6 561	
						−				1	1	0	0	1		
							1	0	0	0	1	0	0	0	136	3
				1	0	0	0	1	0	0	0					
+1	0	0	0	1	0	0	0									
1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	18 496	
						−		1	0	0	1	0	0	0		
										−	1	0	0	0	−8	4
								1	0	0	0	0	0	0	64	5
		1	0	0	0	0	0	0							4 096	
						−				1	0	0	0	0		
									−	1	0	0	0	0	−16	6
						1	0	0	0	0					256	
						−								1		
													−	1	−1	7

En la penúltima columna tenemos 3^{2^n} en base 10, pero esto es sólo para facilitar la lectura de la tabla. Esa columna no es necesaria en los cálculos. El valor de 3^{2^n} está expresado en base 2 en las columnas anteriores.

1. Partimos de $3^{2^0} = 3 = 11_2$.

2. Las tres filas siguientes muestran el producto $11_2 \times 11_2$ calculado con el algoritmo usual de la multiplicación, salvo que no hemos escrito el número dos veces, porque el cálculo se reduce en la práctica a copiar las cifras de 11_2 desplazadas para que las unidades se sitúen debajo de cada 1 y sumar los números resultantes. El resultado es $9 = 1001_2$.

3. Las filas siguientes muestran el producto $1001_2 \times 1001_2$.

4. Las filas siguientes contienen el cálculo de $1010001_2 \times 1010001_2$, y el resultado es 6 561, que en binario se expresa como

   $$11001_2 \times 2^8 + 10100001_2 \equiv -11001_2 + 10100001_2 \pmod{F_3},$$

   donde usamos que $2^8 \equiv -1 \pmod{F_3}$, y las filas siguientes contienen el cálculo de esta resta, que da 136.

   Vemos así que, en general, siempre que llegamos a un número de más de 8 cifras binarias, podemos reducirlo a un número de a lo sumo 8 cifras. Esto asegura que en todo el cálculo nunca será necesario emplear números de más de 16 cifras, pues al elevar al cuadrado un número de a lo sumo 8 cifras obtenemos un número de a lo sumo 16 cifras.

5. Al calcular 136^2 rebasamos también la octava cifra decimal, y ahora en la resta que resulta

   $$1000000_2 - 1001000_2 = -1000_2$$

   el segundo número es mayor, por lo que hay que restarlos al revés y poner un signo negativo que desaparecerá al elevar al cuadrado en el paso siguiente.

6. Procediendo de este modo llegamos a que $3^{2^7} \equiv -1 \pmod{F_3}$, lo que, en virtud del test de Pépin, prueba que F_3 es primo.

Para comprobar que F_6 no es primo se puede calcular análogamente, pero ahora usando números de hasta 128 cifras binarias. Lucas se construyó para ello un tablero dividido en celdas en las que ponía o quitaba piedrecitas, por lo que puede decirse que hizo un cálculo en el sentido etimológico de la palabra.

El resultado final es

$$3^{2^{63}} \equiv 11\,860\,219\,800\,640\,380\,469,$$

que no es $F_6 - 1$, por lo que F_6 resulta ser un número compuesto. Puede probarse que

$$F_6 = 274\,177 \cdot 67\,280\,421\,310\,721.$$

De hecho, el matemático danés Thomas Clausen había encontrado esta descomposición en 1854, pero no la publicó. Fue el matemático francés Fortuné Landry quien en 1880 anunció el resultado, aunque dio pocas indicaciones de cómo lo había obtenido.

Con la ayuda de ordenadores, el test de Pépin se ha aplicado a varios números de Fermat (todos los cuales han resultado ser compuestos), al menos se ha podido aplicar para el estudio de hasta F_{24}.

Capítulo VII

La ley de reciprocidad cuadrática

En los capítulos anteriores hemos observado algunos casos particulares de lo que, vagamente, hemos llamado "reciprocidad cuadrática". Concretamente, hemos visto que para $d = -1, \pm 2, \pm 3$, se cumple que el hecho de que d sea o no un resto cuadrático módulo un primo $q \neq d$, que es una condición que en principio depende del resto de d módulo q, depende también del resto de q módulo $4d$. Concretamente:

$$\left(\frac{-1}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv 1 \text{ (mód 4)}$$

$$\left(\frac{2}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv \pm 1 \text{ (mód 8)}$$

$$\left(\frac{-2}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv 1, 3 \text{ (mód 12)}$$

$$\left(\frac{3}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv \pm 1 \text{ (mód 12)}$$

$$\left(\frac{-3}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv 1, 7 \text{ (mód 12)}.$$

En realidad, para $d = -3$ hemos obtenido $q \equiv 1$ (mód 3), pero (para un número impar q) esto equivale a la condición $q \equiv 1, 7$ (mód 12), y así consideramos siempre restos módulo $4d$. De todos modos, el caso $d < 0$ se reduce al opuesto en virtud del carácter (completamente) multiplicativo del símbolo de Legendre y la primera relación. Por ello podemos limitarnos a considerar valores positivos. Por el mismo motivo podemos restringirnos al caso en que d es un primo $p > 0$, pues si queremos saber, por ejemplo, cuando se cumple $(6/q) = 1$, podemos usar que $(6/q) = (2/q)(3/q)$ y estudiar los dos símbolos de Legendre por separado. Concretamente, vemos que q tiene que cumplir a la vez

$$q \equiv \pm 1 \text{ (mód 8)}, \qquad q \equiv \pm 1 \text{ (mód 12)}$$

y esto se traduce en que

$$\left(\frac{6}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv \pm 1, \pm 5 \text{ (mód 24)}.$$

Así pues, podemos plantearnos si el valor de (p/q) (donde p y q son primos distintos) depende siempre del resto de q módulo $4p$. Como el caso $p = 2$ ya lo conocemos y todo primo impar cumple $(p/2) = 1$, nos restringiremos al caso en que q también es impar.

7.1 Formulación de la ley de reciprocidad

Calculando los cuadrados de $\mathbb{Z}_q$ para los distintos primos impares q podemos ir viendo en qué casos 5 es uno de ellos. El signo de $(5/q)$ es el indicado en la tabla siguiente:

+	11	19	29	31	41	59	61	71	79	89	101	109	131	139	149
−	3	7	13	17	23	37	43	47	53	67	73	83	97	103	107

El patrón es claro: 5 es un resto cuadrático módulo los primos que acaban en 1 o en 9 o, en otros términos, la conjetura es:

$$\left(\frac{5}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv \pm 1 \text{ (mód 5)}.$$

Más aún, si observamos que ± 1 son los cuadrados de U_5, la conjetura es que

$$\left(\frac{5}{q}\right) = \left(\frac{q}{5}\right).$$

Consideremos ahora $p = 7$:

+	3	19	29	31	37	47	53	59	83	103	109	113	131	137	139
−	5	11	13	17	23	41	43	61	67	71	73	79	89	97	101

Vemos que $(7/q)$ no depende del resto de q módulo 7, pues, por ejemplo, $5 \equiv 19$ (mód 7) y $3 \equiv 17$ (mód 7). Sin embargo, si calculamos los restos módulo 28 de cada primo q llegamos a la conjetura

$$\left(\frac{7}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv \pm 1, \pm 3, \pm 9 \text{ (mód 28)}.$$

Ejercicio: Calcular los cuadrados en U_{28}.

Veamos $p = 11$:

+	5	7	19	37	43	53	79	83	89	97	107	113	127	131	137
−	3	13	17	23	29	31	41	47	59	61	67	71	73	101	103

Nuevamente vemos que $(11/q)$ no depende del resto de q módulo 11, pues

$$7 \equiv 29 \text{ (mód 11)}, \quad 31 \equiv 53 \text{ (mód 11)}, \quad 37 \equiv 59 \text{ (mód 11)}.$$

En cambio, módulo 44 obtenemos una conjetura coherente:

$$\left(\frac{11}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv \pm 1, \pm 5, \pm 9, \pm 11, \pm 19 \text{ (mód 44)}.$$

Ejercicio: Calcular los cuadrados en U_{44}.

Finalmente, para $p = 13$ tenemos:

$+$	3	17	23	29	43	53	61	79	101	103	107	113	127	131	139
$-$	5	7	11	19	31	37	41	47	59	67	71	73	83	89	97

Calculando los restos módulo 13 obtenemos la conjetura

$$\left(\frac{13}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv \pm 1, \pm 3, \pm 4 \text{ (mód 13)},$$

pero estas clases son los cuadrados módulo 13, por lo que la conjetura es equivalente a

$$\left(\frac{13}{q}\right) = \left(\frac{q}{13}\right).$$

Vemos, pues, que para algunos primos p se cumple que

$$\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right),$$

mientras que, para otros, (p/q) depende del resto de q módulo $4p$.

Ejercicio: Para $p = 5, 13$, expresar la conjetura en términos del resto de q módulo 20 y 52 (respectivamente) y calcular los cuadrados de U_{20} y U_{52}, respectivamente. Formular una conjetura válida para $p = 3, 5, 7, 11, 13$.

Para tratar de distinguir cuándo se da el caso $(p/q) = (q/p)$, la tabla siguiente muestra el valor de

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right),$$

así la igualdad indicada se da cuando en una fila todos los signos son positivos:

	3	5	7	11	13	17	19	23	29	31	37	41	43	47	53
3	0	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$	$-$	$+$
5	$+$	0	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
7	$-$	$+$	0	$-$	$+$	$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$	$-$	$+$
11	$-$	$+$	$-$	0	$+$	$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$	$-$	$+$
13	$+$	$+$	$+$	$+$	0	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
17	$+$	$+$	$+$	$+$	$+$	0	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
19	$-$	$+$	$-$	$-$	$+$	$+$	0	$-$	$+$	$-$	$+$	$+$	$-$	$-$	$+$
23	$-$	$+$	$-$	$-$	$+$	$+$	$-$	0	$+$	$-$	$+$	$+$	$-$	$-$	$+$
29	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	0	$+$	$+$	$+$	$+$	$+$	$+$
31	$-$	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$+$	0	$+$	$+$	$-$	$-$	$+$
37	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	0	$+$	$+$	$+$	$+$
41	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	0	$+$	$+$	$+$
43	$-$	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$-$	$+$	$+$	0	$-$	$+$
47	$-$	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$	0	$+$
53	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	0

Vemos que $(p/q) = (q/p)$ se da cuando

$$p = 5, \quad 13, \quad 17, \quad 29, \quad 37, \quad 41, \quad 53, \quad \ldots$$

No es difícil captar la regla general: el primer caso se da si $p \equiv 1 \pmod 4$. Pero la tabla muestra una información adicional sobre lo que sucede cuando $p \equiv -1 \pmod 4$: los unos aparecen en las columnas correspondientes a los primos $q \equiv 1 \pmod 4$.

Equivalentemente, concluimos que la igualdad $(p/q) = (q/p)$ se da exactamente cuando al menos uno de los dos primos es congruente con 1 módulo 4. Esto puede enunciarse así:

Teorema 7.1 (Ley de reciprocidad cuadrática) *Si $p, q > 0$ son dos primos impares distintos, entonces*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}.$$

Leyes suplementarias

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}, \qquad \left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}.$$

Observemos que $(-1)^{(p-1)(q-1)/4} = ((-1)^{(p-1)/2})^{(q-1)/2}$ toma el valor 1 cuando al menos uno de los dos primos p, q es congruente con 1 módulo 4, por lo que esta formulación expresa ciertamente la conjetura a la que hemos llegado. Por otro lado, las leyes suplementarias ya las tenemos demostradas (teoremas 3.38 y 5.11).

La ley de reciprocidad cuadrática aparece implícitamente en ciertas conjeturas de Euler, y el primero en demostrarla fue Gauss, si bien su primera demostración era muy complicada. Posteriormente encontró muchas demostraciones alternativas, y hoy en día es probablemente el teorema del que más demostraciones esencialmente distintas se conocen, desde las más elementales hasta las basadas en los resultados más profundos de la teoría algebraica de números. En la sección siguiente presentaremos dos de estas demostraciones.

Observemos que la ley de reciprocidad explica por qué (p/q) depende del resto de q módulo p si $p \equiv 1 \pmod 4$ y del resto módulo $4p$ en caso contrario. En el primer caso afirma que $(p/q) = (q/p)$, lo cual sólo depende del resto de q módulo p, mientras que en el segundo caso $(p/q) = \pm(q/p)$, donde el signo depende del resto de q módulo 4 y el símbolo de Legendre del resto de q módulo p, luego la expresión completa depende del resto de q módulo $4p$.

Observemos también que con la ley de reciprocidad cuadrática es fácil calcular cualquier símbolo de Legendre. Por ejemplo:

$$\left(\frac{19}{53}\right) = \left(\frac{53}{19}\right) = \left(\frac{15}{19}\right) = \left(\frac{3}{19}\right)\left(\frac{5}{19}\right) =$$

$$-\left(\frac{19}{3}\right)\left(\frac{19}{5}\right) = -\left(\frac{1}{3}\right)\left(\frac{2^2}{5}\right) = -\left(\frac{2}{5}\right)^2 = -1.$$

Por otra parte, con unos pocos cálculos, a partir de los ejemplos que hemos analizado, el lector podría haber llegado a la conjetura siguiente:

Teorema 7.2 *Si $p, q > 0$ son dos primos impares distintos, entonces*

$$\left(\frac{p}{q}\right) = 1 \quad \textit{si y sólo si} \quad q \equiv \pm c^2 \pmod{4p}, \ \textit{para cierto entero impar } c.$$

Euler formuló conjeturas que eran claramente equivalentes a este resultado, si bien nunca llegó a escribirlo explícitamente. Vamos a probar que esto es equivalente a la ley de reciprocidad cuadrática.

Demostración: Llamemos $p^* = (-1)^{(p-1)/2}p$. Entonces, la ley de reciprocidad cuadrática es equivalente a

$$\left(\frac{p^*}{q}\right) = \left(\frac{q}{p}\right).$$

En efecto, si $p \equiv 1 \pmod 4$ tenemos que $p^* = p$ y la igualdad es lo que afirma la ley de reciprocidad en este caso, mientras que si $p \equiv -1 \pmod 4$ entonces $p^* = -p$ y la igualdad es

$$\left(\frac{-1}{q}\right)\left(\frac{p}{q}\right) = (-1)^{(q-1)/2}\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right),$$

que también es lo que afirma la ley de reciprocidad en este caso. A su vez, como ambos miembros valen ± 1, la ley de reciprocidad equivale a que

$$\left(\frac{p^*}{q}\right) = 1 \quad \text{si y sólo si} \quad \left(\frac{q}{p}\right) = 1.$$

Por lo tanto, basta probar que

$$\left(\frac{p^*}{q}\right) = 1 \quad \text{si y sólo si} \quad q \equiv \pm c^2 \pmod{4p}, \ \text{para cierto entero impar } c.$$

En efecto, si $(p^*/q) = 1$, esto significa que $(-1)^{(p-1)/2}p \equiv c^2 \pmod q$, para cierto entero c que, cambiándolo si es preciso por $c + q$, podemos tomar impar. Entonces

$$\begin{aligned} p &\equiv (-1)^{(p-1)/2}c^2 \pmod q \\ p &\equiv (-1)^{(p-1)/2} \equiv (-1)^{(p-1)/2}c^2 \pmod 4 \end{aligned}$$

donde hemos usado que todo cuadrado impar es congruente con 1 módulo 4. De aquí se sigue que

$$p \equiv (-1)^{(p-1)/2}c^2 \pmod{4q}.$$

Recíprocamente, si $p \equiv \pm c^2 \pmod{4q}$, el signo es el mismo que en la congruencia $p \equiv \pm 1 \pmod 4$, luego es $(-1)^{(p-1)/2}$, luego $(-1)^{(p-1)/2}p \equiv c^2 \pmod q$, es decir, $(p^*/q) = 1$. ∎

7.2 La prueba de Rousseau

Veamos una prueba de la ley de reciprocidad publicada por Rousseau en 1991 que, siendo elemental, es también conceptualmente muy simple. Está basada en el teorema chino del resto. Dados dos primos impares distintos $p, q > 0$, éste nos asegura que la aplicación

$$\mathbb{Z}_{pq} \longrightarrow \mathbb{Z}_p \times \mathbb{Z}_q$$

dada por $\bar{n} \mapsto (\bar{n}, \bar{n})$ nos permite identificar ambos anillos (donde la suma y el producto en el segundo de ellos se calcula operando componente a componente). En particular se restringe a una aplicación

$$U_{pq} \longrightarrow U_p \times U_q$$

que nos permite identificar las unidades de cada anillo. Por ejemplo, aquí tenemos la correspondencia para $p = 3$ y $q = 7$:

$-\overline{10}$	$-\bar{8}$	$-\bar{5}$	$-\bar{4}$	$-\bar{2}$	$-\bar{1}$
$(-\bar{1}, -\bar{3})$	$(\bar{1}, -\bar{1})$	$(\bar{1}, \bar{2})$	$(-\bar{1}, \bar{3})$	$(\bar{1}, -\bar{2})$	$(-\bar{1}, -\bar{1})$
$\bar{1}$	$\bar{2}$	$\bar{4}$	$\bar{5}$	$\bar{8}$	$\overline{10}$
$(\bar{1}, \bar{1})$	$(-\bar{1}, \bar{2})$	$(\bar{1}, -\bar{3})$	$(-\bar{1}, -\bar{2})$	$(-\bar{1}, \bar{1})$	$(\bar{1}, \bar{3})$

Es claro que si en U_{pq} tomamos representantes de sus clases entre $-(pq-1)/2$ y $(pq-1)/2$ tenemos que la mitad de las clases son positivas y la otra mitad son negativas. Llamemos $P \subset U_{pq}$ al conjunto de las clases con representante positivo.

Similarmente, llamamos $\tilde{P} \subset U_p \times U_q$ al conjunto de todos los pares cuya segunda componente admite un representante entre 0 y $(q-1)/2$. En nuestro ejemplo:

P	$\bar{1}$	$\bar{2}$	$\bar{4}$	$\bar{5}$	$\bar{8}$	$\overline{10}$
$\tilde{P}$	$(\bar{1}, \bar{1})$	$(-\bar{1}, \bar{2})$	$(-\bar{1}, \bar{3})$	$(\bar{1}, \bar{2})$	$(-\bar{1}, \bar{1})$	$(\bar{1}, \bar{3})$

Es obvio que P y $\tilde{P}$ tienen el mismo número de elementos, pero la relación entre ambos es mucho más estrecha. Los elementos de U_{pq} se dividen en pares $\bar{n}$ y $-\bar{n}$, con $\bar{n}$ en P, y que se corresponden con pares $(\bar{n}, \bar{n})$ y $(-\bar{n}, -\bar{n})$. No es necesariamente cierto que el primero esté en $\tilde{P}$, pero uno (y sólo uno) de los dos lo está. Por lo tanto, definiendo $\epsilon_n = \pm 1$ adecuadamente, podemos hacer que $\epsilon_n(\bar{n}, \bar{n})$ esté en $\tilde{P}$, y así tenemos una correspondencia biunívoca entre los conjuntos P y $\tilde{P}$. En particular, si multiplicamos todos los pares $(\bar{n}, \bar{n})$ con $\bar{n}$ en P, obtendremos el producto de todos los elementos de $\tilde{P}$ salvo a lo sumo un signo (el producto de todos los ϵ_n), es decir:

$$\prod_{(\bar{a},\bar{b})\in\tilde{P}} (\bar{a}, \bar{b}) = \epsilon \prod_{\bar{n}\in P} (\bar{n}, \bar{n}),$$

donde $\epsilon = \pm 1$. Vamos a desarrollar ambos miembros. Para ello llamamos $p_0 = (p-1)/2$ y $q_0 = (q-1)/2$. Así, cuando el par $(\bar{a}, \bar{b})$ recorre $\tilde{P}$, tenemos que

a recorre todos los números entre 1 y $p-1$ (y cada número aparece como primera componente de q_0 pares), mientras que b recorre todos los números entre 1 y q_0 (y cada número aparece $p-1 = 2p_0$ veces). Por lo tanto:

$$\prod_{(\bar{a},\bar{b})\in\tilde{P}} (\bar{a},\bar{b}) = (\overline{(p-1)!}^{q_0}, \overline{q_0!}^{2p_0}).$$

Notemos que $\overline{(q-1)!}$ es el producto de todos los elementos de U_q, pero, tomando representantes de las clases entre $-q_0$ y q_0, el producto de las positivas es $\overline{q_0!}$ y el de las negativas $(-1)^{q_0}\overline{q_0!}$, luego $(q-1)! \equiv (q_0!)^2(-1)^{q_0}$ (mód q). Usando esto y el teorema de Wilson 3.10:

$$\prod_{(\bar{a},\bar{b})\in\tilde{P}} (\bar{a},\bar{b}) = (\overline{(p-1)!}^{q_0}, (\overline{(q-1)!}(-1)^{q_0})^{p_0})) = ((-\bar{1})^{q_0}, (-\bar{1})^{p_0}(-\bar{1})^{p_0q_0}).$$

Ahora vamos a desarrollar el producto $\prod_{\bar{n}\in P} (\bar{n},\bar{n})$. Consideramos su primera componente, de modo que las clases son módulo p. Se obtiene multiplicando las clases de todos los números $1 \le n \le (pq-1)/2$ tales que $(n,pq) = 1$. Esto es lo mismo que si multiplicamos por todas las clases con $p \nmid n$ y luego dividimos entre las clases con $q \mid n$:

$$\prod_{\bar{n}\in P} \bar{n} = \prod_{p\nmid n} \bar{n} \left(\prod_{q|n} \bar{n}\right)^{-1}.$$

Los números $1 \le n \le (pq-1)/2 = q_0p + p_0 = p_0q + q_0$ tales que $p \nmid n$ se pueden agrupar así:

$$\begin{array}{rrcr}
1, & 2, & \ldots, & p-1,\\
p+1, & p+2, & \ldots, & 2p-1,\\
\vdots & \vdots & & \vdots\\
(q_0-1)p+1, & (q_0-1)p+2, & \ldots, & q_0p-1\\
q_0p+1, & \ldots, & q_0p+p_0 &
\end{array}$$

Y al tomar clases módulo p, el producto de cada fila es $\overline{(p-1)!}$ salvo en el caso de la última, que es $\overline{p_0!}$. Por otro lado,

$$\prod_{q|n} \bar{n} = \bar{q}\cdot\bar{2}\bar{q}\cdot\bar{3}\bar{q}\cdots\bar{p}_0\bar{q} = \overline{(p_0!)}\,\bar{q}^{p_0}.$$

En total (y aplicando de nuevo el teorema de Wilson) queda que

$$\prod_{\bar{n}\in P} \bar{n} = \overline{(p-1)!}^{q_0}\ \overline{p_0!}\ (\overline{(p_0!)}\,\bar{q}^{p_0})^{-1} = (-\bar{1})^{q_0}\bar{q}^{-(p-1)/2} = (-\bar{1})^{q_0}\left(\frac{q}{p}\right),$$

donde al final hemos aplicado el criterio de Euler 5.8. La segunda componente se desarrolla del mismo modo y en total obtenemos que

$$((-\bar{1})^{q_0}, (-\bar{1})^{p_0}(-\bar{1})^{p_0q_0}) = \epsilon((-\bar{1})^{q_0}\left(\frac{q}{p}\right), (-\bar{1})^{p_0}\left(\frac{p}{q}\right)).$$

Igualando las componentes queda

$$\epsilon\left(\frac{q}{p}\right) \equiv 1 \text{ (mód } p), \qquad \epsilon\left(\frac{p}{q}\right) \equiv (-1)^{(p-1)(q-1)/4} \text{ (mód } q).$$

Como todos los términos son ± 1, las congruencias son igualdades:

$$\epsilon = \left(\frac{q}{p}\right), \qquad \epsilon\left(\frac{p}{q}\right) = (-1)^{(p-1)(q-1)/4},$$

de donde obtenemos la ley de reciprocidad. ∎

7.3 Sumas de Gauss

En los capítulos anteriores hemos probado algunos casos particulares de la ley de reciprocidad cuadrática relacionando $\sqrt{2}$ y $\sqrt{-3}$ con raíces primitivas de la unidad. Vamos a ver que esa línea argumental nos da una prueba alternativa de la ley de reciprocidad cuadrática.

Sea $p > 0$ un primo impar y sea k un cuerpo de característica distinta de p. En $k[x]$ consideramos el polinomio

$$x^p - 1 = (x-1)(x^{p-2} + x^{p-3} + \cdots + x + 1).$$

El segundo factor no tiene por qué ser irreducible en $k[x]$, pero podemos tomar un factor irreducible y, por el teorema 6.1 obtenemos un cuerpo K que contiene a k como subcuerpo en el que dicho factor tiene una raíz ζ. Esta raíz cumple $\zeta^p = 1$, pero $\zeta \neq 1$, pues

$$1^{p-2} + 1^{p-3} + \cdots + 1 + 1 = p \neq 0.$$

Así pues, ζ es una raíz p-ésima de la unidad, que puede estar en k o no.

Definimos la *suma de Gauss*

$$G(p) = \left(\frac{1}{p}\right)\zeta + \left(\frac{2}{p}\right)\zeta^2 + \cdots + \left(\frac{p-1}{p}\right)\zeta^{p-1}.$$

El interés de esta definición radica en el hecho siguiente:

$$G(p)^2 = (-1)^{(p-1)/2}p. \tag{7.1}$$

En efecto, tenemos que

$$G(p)^2 = \sum_{a,b=1}^{p-1}\left(\frac{a}{p}\right)\left(\frac{b}{p}\right)\zeta^{a+b}.$$

Ahora bien, cada sumando depende únicamente del resto de b módulo p, por lo que si cambiamos b por ab, la suma no cambia, ya que ab recorre también todas las clases no nulas módulo p. Por consiguiente:

$$G(p)^2 = \sum_{a,b=1}^{p-1} \left(\frac{a}{p}\right)\left(\frac{ab}{p}\right)\zeta^{a+ab} = \sum_{a,b=1}^{p-1} \left(\frac{b}{p}\right)\zeta^{a+ab} = \sum_{b=1}^{p-1}\Big(\sum_{a=1}^{p-1}\zeta^{a(1+b)}\Big)\left(\frac{b}{p}\right).$$

Si $1+b \not\equiv 0 \pmod{p}$, entonces $a(1+b)$ recorre todos los restos no nulos módulo p cuando varía a, por lo que la suma de las potencias $\zeta^{a(1+b)}$ es

$$\zeta + \cdots + \zeta^{p-1} = -1.$$

Si $1+b \equiv 0 \pmod{p}$, entonces $\zeta^{a(1+b)} = 1$ y la suma es $p-1$, luego

$$G(p)^2 = -\sum_{b=1}^{p-2}\left(\frac{b}{p}\right) + (p-1)\left(\frac{-1}{p}\right) = -\sum_{b=1}^{p-1}\left(\frac{b}{p}\right) + p\left(\frac{-1}{p}\right).$$

Ahora bien, como hay el mismo número de restos cuadráticos que de restos no cuadráticos, la suma es nula, y queda la igualdad que queríamos probar. ∎

Una de las pruebas de Gauss de la ley de reciprocidad cuadrática usaba sumas de Gauss sobre el cuerpo $\mathbb{Q}$ de los números racionales, pero el argumento se simplifica si consideramos $G(p)$ definida sobre $\mathbb{Z}_q$, donde p y q son primos impares distintos.

En efecto, en la prueba de que el teorema 7.2 equivale a la ley de reciprocidad cuadrática hemos visto que ésta equivale a que

$$\left(\frac{p^*}{q}\right) = \left(\frac{q}{p}\right),$$

donde $p^* = (-1)^{(p-1)/2}p$. Acabamos de ver que $G(p)$ es una raíz cuadrada de p^* en un cuerpo que contiene a $\mathbb{Z}_q$. Como el polinomio $x^2 - p^*$ sólo puede tener dos raíces en dicho cuerpo, se cumplirá que

$$\left(\frac{p^*}{q}\right) = 1 \quad \text{si y sólo si} \quad G(p) \text{ está en } \mathbb{Z}_q.$$

Pero $\mathbb{Z}_q$ está formado por las raíces del polinomio $x^p - x$, luego la suma de Gauss $G(p)$ está en $\mathbb{Z}_q$ si y sólo si $G(p)^q = G(p)$. Ahora bien:

$$G(p)^q = \sum_{a=1}^{p-1}\left(\frac{a}{p}\right)\zeta^{aq} = \left(\frac{q}{p}\right)\sum_{a=1}^{p-1}\left(\frac{aq}{p}\right)\zeta^{aq},$$

pero aq recorre todas las clases de restos no nulas módulo p cuando varía a, luego

$$G(p)^q = \left(\frac{q}{p}\right)\sum_{a=1}^{p-1}\left(\frac{a}{p}\right)\zeta^{a} = \left(\frac{q}{p}\right)G(p).$$

Esto implica que

$$\left(\frac{p^*}{q}\right) = \left(\frac{q}{p}\right).$$

∎

Definición alternativa En realidad Gauss definió las que hoy se conocen como sumas de Gauss como

$$G(p) = \sum_{x=1}^{p} \zeta^{x^2}. \tag{7.2}$$

La equivalencia con la definición que hemos dado es fácil de probar: cuando x varía de 1 a $p-1$, tenemos que x^2 recorre dos veces cada resto cuadrático módulo p, luego, con esta definición.

$$G(p) = \sum_{a=1}^{p} \left(\left(\frac{a}{p}\right) + 1\right) \zeta^a,$$

pues en esta expresión las potencias de ζ cuyo exponente es un resto cuadrático (no nulo) aparecen dos veces, las que tienen exponente resto no cuadrático no aparecen y $\zeta^{p^2} = 1$ aparece una vez. Si ahora tenemos en cuenta que

$$\sum_{a=1}^{p} \zeta^a = 0,$$

concluimos que $G(p)$ coincide con la suma que habíamos definido antes.

Con esta expresión es muy fácil probar que, en $\mathbb{Z}_q$, se cumple que

$$G(p)^q = \left(\frac{q}{p}\right) G(p),$$

pues en principio

$$G(p)^q = \sum_{x=1}^{p} \zeta^{qx^2},$$

luego si $(q/p) = 1$ entonces $q \equiv u^2$ (mód p), por lo que $qx^2 \equiv (ux)^2$ (mód p) y cuando x recorre las clases de $\mathbb{Z}_p$ lo mismo hace ux, por lo que $G(p)^q = G(p)$. En cambio, si $(q/p) = -1$, entonces los exponentes de ζ en $G(p)^q$ recorren los dos veces cada resto no cuadrático módulo p (más el 0 una vez), mientras que los exponentes en $G(p)$ recorren dos veces cada resto cuadrático no nulo módulo p (más el 0 una vez), luego

$$G(p) + G(p)^q = 2(1 + \zeta + \cdots \zeta^{p-1}) = 0.$$

Por otro lado, la prueba de (7.1) usando la definición de Gauss también es sencilla si usamos una caso particular del teorema 5.12. En efecto, en principio

$$G(p)^2 = \sum_{x,y=1}^{p} \zeta^{x^2+y^2},$$

pero, según, 5.12, sabemos que $x^2 + y^2$ toma $p - (-1/p)$ veces cada valor de U_p y $p + (p-1)(-1/p)$ veces el valor 0, luego

$$\begin{aligned} G(p)^2 &= p + (p-1)\left(\tfrac{-1}{p}\right) + \left(p - \left(\tfrac{-1}{p}\right)\right) \sum_{z=1}^{p-1} \zeta^z \\ &= p + (p-1)\left(\tfrac{-1}{p}\right) - p + \left(\tfrac{-1}{p}\right) = (-1)^{(p-1)/2}\, p. \end{aligned}$$

Esto nos da una variante de la prueba de la ley de reciprocidad cuadrática.

7.4 Restos cuadráticos generales

Hasta aquí hemos considerado restos cuadráticos módulo números primos, pero este caso no dista mucho del caso general.

Teorema 7.3 *Si p es un primo impar y $(a,p) = 1$, entonces a es un resto cuadrático módulo p^k si y sólo si es un resto cuadrático módulo p.*

DEMOSTRACIÓN: Vamos a probar inductivamente que, si $k \geq 1$, se cumple que a es un resto cuadrático módulo p^k si y sólo si lo es módulo p^{k+1}. Una implicación es obvia. Supongamos que

$$a = r^2 + mp^k.$$

Entonces

$$(r + np^k)^2 = r^2 + 2rnp^k + n^2p^{2k} \equiv a + (2rn - m)p^k \pmod{p^k}.$$

Si elegimos n tal que $2rn \equiv m \pmod{p}$, se cumple que $a \equiv (r+np^k)^2 \pmod{p^{k+1}}$. Notemos que la congruencia $2rn \equiv m \pmod{p}$ siempre tiene solución, pues $p \nmid 2r$. ∎

Para $p = 2$ la situación es ligeramente distinta. Todo número impar a es un resto cuadrático módulo 2, es un resto cuadrático módulo 4 si y sólo si cumple $a \equiv 1 \pmod{4}$ y, en general:

Teorema 7.4 *Un número impar a es un resto cuadrático módulo 2^k con $k \geq 3$ si y sólo si $a \equiv 1 \pmod{8}$.*

DEMOSTRACIÓN: Veamos igualmente que a es un resto cuadrático módulo 2^k si y sólo si lo es módulo 2^{k+1}, pero ahora la prueba requiere suponer que $k \geq 3$. En efecto, si

$$a = r^2 + m2^k,$$

entonces

$$(r + m2^{k-1})^2 = r^2 + m2^k + m^2 2^{2k-2} \equiv a \pmod{2^{k+1}},$$

donde usamos que $2k - 2 \geq k + 1$ (porque $k \geq 3$). Ahora bien, es fácil ver que un número impar a es un resto cuadrático módulo 8 si y sólo si $a \equiv 1 \pmod{8}$. ∎

Por último, el teorema chino del resto implica inmediatamente:

Teorema 7.5 *Si $n = p_1^{e_1} \cdots p_r^{e_r}$ es la descomposición en factores primos de un número natural $n > 1$, entonces un entero a primo con n es un resto cuadrático módulo n si y sólo si lo es módulo $p_i^{e_i}$ para todo i.*

DEMOSTRACIÓN: Si $a \equiv r_i^2 \pmod{p_i^{e_i}}$ para todo i, el teorema chino del resto nos asegura que existe un entero r tal que $r \equiv r_i \pmod{p_i^{e_i}}$ para todo i, luego $a \equiv r^2 \pmod{p_i^{e_i}}$, luego $p_i^{e_i} \mid a - r^2$, luego $n \mid a - r^2$, es decir, $a \equiv r^2 \pmod{n}$. ∎

Ejemplo Lo que muestra la demostración del teorema 7.3 es que una raíz cuadrada de a módulo p^{k+1} puede obtenerse sumando a una raíz cuadrada módulo p^k un término de la forma np^{k+1}.

Por ejemplo, si tomamos $p = 5$, $a = 6$ y elegimos una raíz cuadrada de a módulo 5 (hay dos $r_1 = \pm 1$) por ejemplo, $r_1 = 1$, tenemos que

$$6 = 1^2 + 1 \cdot 5,$$

de modo que, con la notación de la demostración, tenemos que $m = 1$. Entonces podemos calcular $n = 3$ (notemos que n siempre puede elegirse entre 0 y $p-1$, o en cualquier otro conjunto de representantes de las clases de $\mathbb{Z}_p$). Esto significa que una raíz cuadrada de 6 módulo 5^2 es

$$r_2 = 1 + 3 \cdot 5.$$

De hecho, $6 = r_2^2 - 2 \cdot 5^3$, por lo que ahora $m = -10$, lo que nos lleva a que $n = 0$, es decir, que

$$r_3 = 1 + 3 \cdot 5 + 0 \cdot 5^2$$

es una raíz cuadrada de 6 módulo 5^3, con $m = -2$. Esto nos da $n = 4$, luego

$$r_4 = 1 + 3 \cdot 5 + 0 \cdot 5^2 + 4 \cdot 5^3$$

es una raíz cuadrada de 6 módulo 5^4. Así podemos ir generando una serie de potencias

$$\sqrt{6} = 1 + 3 \cdot 5 + 0 \cdot 5^2 + 4 \cdot 5^3 + 2 \cdot 5^4 + 1 \cdot 5^5 + 2 \cdot 5^6 + 3 \cdot 5^7 + \cdots$$

que truncada hasta la potencia 5^k nos da una raíz cuadrada de 6 módulo 5^{k+1}.

Si empezamos con la raíz $r_1 = 4$ módulo 5 obtenemos

$$\sqrt{6} = 4 + 1 \cdot 5 + 4 \cdot 5^2 + 0 \cdot 5^3 + 2 \cdot 5^4 + 3 \cdot 5^5 + 2 \cdot 5^6 + 1 \cdot 5^7 + \cdots$$

Nota Si el lector tiene curiosidad por estas series de potencias, puede buscar información sobre los llamados números p-ádicos, pues sucede que las series que hemos calculado convergen realmente a raíces cuadradas de 6 en el cuerpo $\mathbb{Q}_5$ de los números pentádicos. ∎

Ejercicio: Calcular los primeros términos del desarrollo de las dos raíces cuadradas de 7 en series de potencias de 3 con coeficientes $-1, 0, 1$.

7.5 El símbolo de Jacobi

El símbolo de Legendre (n/p) sólo está definido cuando $p > 0$ es primo. Sin embargo, Jacobi se dio cuenta de que es posible operar más cómodamente con símbolos de Legendre si extendemos la definición para eliminar (casi completamente) esta restricción:

Definición 7.6 Si n y m son números enteros y $m > 0$ es impar, de modo que $m = p_1 \cdots p_s$ es su descomposición en factores primos, definimos el *símbolo de Jacobi*

$$\left(\frac{n}{m}\right) = \left(\frac{n}{p_1}\right) \cdots \left(\frac{n}{p_s}\right),$$

entendiendo que $(m/1) = 1$.

Naturalmente, los símbolos que aparecen en el miembro derecho de la definición son símbolos de Legendre. Se usa la misma notación porque, evidentemente, si m es un primo impar, el símbolo de Jacobi (n/m) coincide con el símbolo de Legendre.

Veamos las propiedades básicas. La mayor parte de ellas son consecuencia inmediata de las propiedades correspondientes del símbolo de Legendre y de la definición del símbolo de Jacobi.

1. $\left(\frac{n}{m}\right) = \begin{cases} \pm 1 & \text{si } (n,m) = 1, \\ 0 & \text{si } (n,m) \neq 1. \end{cases}$

2. $\left(\frac{n_1 n_2}{m}\right) = \left(\frac{n_1}{m}\right)\left(\frac{n_2}{m}\right), \quad \left(\frac{n}{m_1 m_2}\right) = \left(\frac{n}{m_1}\right)\left(\frac{n}{m_2}\right).$

3. Si $n_1 \equiv n_2 \pmod{m}$, entonces $\left(\frac{n_1}{m}\right) = \left(\frac{n_2}{m}\right)$.

4. Si m y n son impares, positivos y $(m,n) = 1$, entonces

$$\left(\frac{m}{n}\right)\left(\frac{m}{n}\right) = (-1)^{(m-1)(n-1)/4}.$$

5. $\left(\frac{-1}{n}\right) = (-1)^{(n-1)/2}, \qquad \left(\frac{2}{n}\right) = (-1)^{(n^2-1)/8}.$

6. Si $n_1 \equiv n_2 \pmod{4m}$, entonces

$$\left(\frac{m}{n_1}\right) = \left(\frac{m}{n_2}\right).$$

   Si $m \equiv 0, 1 \pmod 4$, basta exigir que $n_1 \equiv n_2 \pmod m$.

Las tres primeras propiedades son inmediatas. Para probar la cuarta observamos que se cumple trivialmente si $m = 1$. Supongamos ahora que m es un primo impar. Si desarrollamos (m/n) según los factores primos de n, invertimos cada símbolo de Legendre por la ley de reciprocidad cuadrática y volvemos a agrupar los factores, llegamos a una igualdad de la forma $(m/n) = \pm(n/m)$. Tenemos que analizar el signo.

Si $m \equiv 1 \pmod 4$, entonces la ley de reciprocidad cuadrática para los factores primos de n es $(m/p) = (p/m)$, luego la igualdad final se cumple con signo positivo.

Si $m \equiv -1$ (mód 4), entonces $(m/p) = \pm(p/m)$ según si $p \equiv \pm 1$ (mód 4), pero el número de divisores primos de p que cumplen $p \equiv -1$ (mód 4) será par o impar según si $n \equiv 1$ (mód 4) o bien $n \equiv -1$ (mód 4), y esto se traducirá a su vez en que el signo final será positivo o negativo, respectivamente. En suma, llegamos a que $(m/n)(n/m) = (-1)^{(m-1)(n-1)/4}$.

Esto prueba la relación cuando m es primo (y n es arbitrario). En el caso general, desarrollamos (m/n) en producto de los símbolos de Jacobi correspondientes a los factores primos de m, aplicamos la parte ya probada en lugar de la ley de reciprocidad cuadrática para el símbolo de Legendre y volvemos a agrupar. El mismo razonamiento anterior nos da ahora la relación para m arbitrario.

Para la quinta propiedad observamos que si desarrollamos $(-1/n)$ según la definición del símbolo de Jacobi en producto de símbolos de Legendre, el resultado será 1 si y sólo si el número de divisores primos de n que cumplen $p \equiv -1$ (mód 4) es par, lo cual equivale a que $n \equiv 1$ (mód 4).

Similarmente, tendremos que $(2/n) = 1$ si y sólo si el número de divisores primos de n que cumplen $p \equiv \pm 3, 5$ (mód 8) es par, lo cual equivale a que $n \equiv \pm 1$ (mód 8) (pues $3^2 \equiv 5^2 \equiv 1$ (mód 8), $3 \cdot 5 \equiv -1$ (mód 8)).

Para la sexta propiedad descomponemos $m = \epsilon 2^j m'$, donde m' es impar y $\epsilon = \pm 1$. Entonces

$$\begin{aligned}\left(\frac{m}{n_1}\right) &= \left(\frac{\epsilon}{n_1}\right)\left(\frac{2}{n_1}\right)^j\left(\frac{m'}{n_1}\right) = \left(\frac{\epsilon}{n_1}\right)\left(\frac{2}{n_1}\right)^j (-1)^{(m'-1)(n_1-1)/4}\left(\frac{n_1}{m'}\right),\\ \left(\frac{m}{n_2}\right) &= \left(\frac{\epsilon}{n_2}\right)\left(\frac{2}{n_2}\right)^j\left(\frac{m'}{n_2}\right) = \left(\frac{\epsilon}{n_2}\right)\left(\frac{2}{n_2}\right)^j (-1)^{(m'-1)(n_2-1)/4}\left(\frac{n_2}{m'}\right).\end{aligned}$$

Tanto si $n_1 \equiv n_2$ (mód $4m$) como si $n_1 \equiv n_2$ (mód m) y $m \equiv 0$ (mód 4), tenemos que $n_1 \equiv n_2$ (mód m)$'$ y $n_1 \equiv n_2$ (mód 4). La primera de estas dos congruencias implica que los últimos símbolos de Jacobi son iguales. Por la segunda, los demás factores de las últimas expresiones también son iguales, salvo a lo sumo $(2/n_i)^j$. Si j es par ambos son iguales a 1 y si j es impar entonces bajo la hipótesis $n_1 \equiv n_2$ (mód $4m$) concluimos que m es par, luego $n_1 \equiv n_2$ (mód 8) y, bajo la hipótesis $m \equiv 0$ (mód 4) concluimos que $j \geq 3$, luego igualmente $n_1 \equiv n_2$ (mód 8) , luego también se da la igualdad $(2/n_1) = (2/n_2)$, luego $(m/n_1) = (m/n_2)$.

Si $m \equiv 1$ (mód 4) y $n_1 \equiv n_2$ (mód m), entonces $j = 0$. Si $\epsilon = 1$ tenemos simplemente que

$$\left(\frac{m}{n_1}\right) = \left(\frac{n_1}{m}\right) = \left(\frac{n_2}{m}\right) = \left(\frac{m}{n_2}\right).$$

Suponemos, pues, que $\epsilon = -1$, en cuyo caso $m' \equiv -1$ (mód 4), luego

$$(-1)^{(m'-1)(n_1-1)/4} = (-1)^{(n_1-1)/2} = \left(\frac{\epsilon}{n_1}\right),$$

e igualmente con n_2, luego

$$\left(\frac{m}{n_1}\right) = \left(\frac{n_1}{m'}\right) = \left(\frac{n_2}{m'}\right) = \left(\frac{m}{n_2}\right). \qquad \blacksquare$$

Nota Observemos que, para $(m,n) = 1$, no es cierto que

$$\left(\frac{m}{n}\right) = 1 \quad \text{si y sólo si} \quad m \text{ es un resto cuadrático módulo } n.$$

Por ejemplo,

$$\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = (-1)(-1) = 1,$$

pero el hecho de que 2 no sea un resto cuadrático módulo 3 ni módulo 5 implica que tampoco lo es módulo 15.

Lo que sí que es cierto es que si m es un resto cuadrático módulo n, entonces lo es módulo todos los divisores primos de n, luego $(m/n) = 1$ o, recíprocamente, si $(m/n) = -1$, entonces m es un resto no cuadrático módulo n. $\blacksquare$

Aunque el símbolo de Jacobi también tiene interés teórico, una de sus aplicaciones prácticas es que simplifica drásticamente el cálculo de símbolos de Legendre, ya que no necesitamos descomponer en factores primos los "numeradores" para aplicar la ley de reciprocidad cuadrática:

$$\left(\frac{19}{53}\right) = \left(\frac{53}{19}\right) = \left(\frac{15}{19}\right) = -\left(\frac{19}{15}\right) = -\left(\frac{2}{15}\right)^2 = -1.$$

A lo sumo puede hacer falta separar las potencias de 2 que aparezcan en los "numeradores", pero eso es mucho más sencillo que descomponer en factores primos.

7.6 El teorema de Dirichlet

El nombre de "ley de reciprocidad cuadrática" fue acuñado por Legendre, quien en 1785 intentó demostrarla a partir del teorema 3.11. Para ello descompuso el enunciado en ocho casos, según si los primos p y q son congruentes con 1 o con -1 módulo 4 y según el valor de (q/p):

1. Si $p \equiv 1 \pmod 4$, $q \equiv -1 \pmod 4$, $(q/p) = 1$, entonces $(p/q) = 1$,
2. Si $p \equiv -1 \pmod 4$, $q \equiv 1 \pmod 4$, $(q/p) = -1$, entonces $(p/q) = -1$,
3. Si $p \equiv 1 \pmod 4$, $q \equiv -1 \pmod 4$, $(q/p) = -1$, entonces $(p/q) = -1$,
4. Si $p \equiv -1 \pmod 4$, $q \equiv 1 \pmod 4$, $(q/p) = 1$, entonces $(p/q) = 1$,
5. Si $p \equiv 1 \pmod 4$, $q \equiv 1 \pmod 4$, $(q/p) = 1$, entonces $(p/q) = 1$,
6. Si $p \equiv 1 \pmod 4$, $q \equiv 1 \pmod 4$, $(q/p) = -1$, entonces $(p/q) = -1$,
7. Si $p \equiv -1 \pmod 4$, $q \equiv -1 \pmod 4$, $(q/p) = 1$, entonces $(p/q) = -1$,
8. Si $p \equiv -1 \pmod 4$, $q \equiv -1 \pmod 4$, $(q/p) = -1$, entonces $(p/q) = 1$.

(Fue precisamente para abreviar la notación para tratar todos estos casos por lo que introdujo lo que hoy se conoce como símbolo de Legendre.)

Notemos que 2 se sigue de 1 por reducción al absurdo, al igual que 4 de 3 y 6 de 5, por lo que sólo hay que demostrar 1, 3, 5, 7, 8.

Por ejemplo, para probar 1 observamos que si fuera

$$p \equiv 1 \text{ (mód 4)}, \quad q \equiv -1 \text{ (mód 4)}, \quad (q/p) = 1, \quad (p/q) = -1,$$

por 3.11, la ecuación $x^2 + py^2 - qz^2 = 0$ tendría una solución no trivial, pues

$$\left(\frac{-p}{q}\right) = \left(\frac{-1}{q}\right)\left(\frac{p}{q}\right) = 1, \quad \left(\frac{q}{p}\right) = 1$$

y obviamente $-pq$ es un cuadrado módulo 1. Dividiendo entre el máximo común divisor de x, y, z, podemos suponer que éste es 1, con lo que al menos dos de estos valores son impares, pero módulo 4 la ecuación es

$$x^2 + y^2 + z^2 \equiv 0 \text{ (mód 4)},$$

lo cual es imposible, porque los cuadrados módulo 4 son 0, 1 y tiene que haber al menos dos unos.

El caso 7 se razona análogamente, con la ecuación $x^2 - py^2 - qz^2 = 0$. Sin embargo, los casos 1, 2, 7 fueron los únicos que Legendre pudo demostrar completamente. El lector puede tratar de aplicar la misma técnica al caso 8, pero verá que no funciona. No obstante, podemos considerar una variante:

Supongamos que

$$p \equiv -1 \text{ (mód 4)}, \quad q \equiv -1 \text{ (mód 4)}, \quad (q/p) = -1, \quad (p/q) = -1,$$

En principio, tendríamos que considerar la ecuación $x^2 - py^2 - qz^2 = 0$, pero no cumple las hipótesis del teorema 3.11. No obstante, esto se arregla si suponemos que existe un tercer primo $r \equiv 1$ (mód 4) que cumpla

$$\left(\frac{r}{p}\right) = \left(\frac{r}{q}\right) = -1$$

y tomando la ecuación $rx^2 - py^2 - qz^2 = 0$. Ésta sí que cumple las hipótesis del teorema 3.11 y nos lleva a la misma contradicción anterior. Ahora bien, ¿podemos asegurar la existencia de tal primo r?

Podemos fijar restos no cuadráticos u, v módulo p y q respectivamente, y el teorema chino del resto nos asegura que existe un entero m tal que

$$m \equiv 1 \text{ (mód 4)}, \quad m \equiv u \text{ (mód } p), \quad m \equiv v \text{ (mód } q),$$

y la cuestión es si existe un primo r que cumpla $r \equiv m$ (mód $4pq$). Legendre no pudo demostrar esto, pero conjeturó que era cierto. Más en general, conjeturó que, si $(m, n) = 1$, siempre existen primos de la forma $m+nk$, para $k = 0, 1, 2, \ldots$ Euler había conjeturado este hecho para $m = 1$ en 1775, pero Legendre fue el primero en plantear la conjetura general. En términos modernos, equivale a que todas las clases de restos de U_n contienen números primos.

En sus *Disquisitiones arithmeticae,* de 1801, Gauss examinó el argumento de Legendre y concedió la plausibilidad de su conjetura, pero tampoco pudo demostrarla. Ésta fue probada por primera vez por Dirichlet en 1837:

Teorema (Dirichlet) *Toda progresión aritmética $m+nk$, para $k = 0, 1, 2, \ldots$ con $(m, n) = 1$ contiene infinitos primos.*

La prueba de Dirichlet se basaba en la teoría de funciones holomorfas (funciones derivables de variable compleja). No se conoce ninguna prueba puramente algebraica de este resultado. En [ITAn 7.23] damos una de las pruebas más elementales posibles, que únicamente utiliza los conceptos de límite y continuidad.

Sin embargo, el teorema de Dirichlet no es suficiente para demostrar los casos 3 y 5 que requería Legendre para completar su demostración de la ley de reciprocidad. En ellos tuvo que suponer un hecho adicional, a saber, que para cada primo $p \equiv 1 \pmod 4$, existe otro primo $q \equiv -1 \pmod 4$ tal que $(p/q) = -1$. Por desgracia, esto no puede demostrarse sin suponer algún caso de la ley de reciprocidad, lo que invalida el intento de Legendre.[1]

En la sección siguiente presentamos una aplicación interesante de la ley de reciprocidad cuadrática y el teorema de Dirichlet.

7.7 Sumas de tres cuadrados

El problema de determinar qué números naturales pueden expresarse como suma de tres cuadrados es más complicado que los casos de dos y cuatro cuadrados que ya hemos tratado. La tabla siguiente muestra los primeros:

1	2	3	4	5	6	8	9	10	11	12	13	14	16	17	18	19
20	21	22	24	25	26	27	29	30	32	33	34	35	36	37	38	40
41	42	43	44	45	46	48	49	50	51	52	53	54	56	57	58	59
61	62	64	65	66	67	68	69	70	72	73	74	75	76	77	78	80
81	82	83	84	85	86	88	89	90	91	93	94	96	97	98	99	100

Vemos que son muy abundantes. Puestos a formular una conjetura, es preferible fijarse en los números que no son suma de tres cuadrados:

7	15	23	28	31	39	47	55	60	63	71	79	87
92	95	103	111	112	119	124	127	135	143	151	156	159
167	175	183	188	191	199	207	215	220	223	231	239	240
247	252	255	263	271	279	284	287	295	303	311	316	319
327	335	343	348	35	359	367	368	375	380	383	391	399
407	412	415	423	431	439	444	447	448	455	463	471	476

Dejamos que el lector conjeture a partir de aquí el resultado que vamos a demostrar. De momento probamos lo siguiente:

Teorema 7.7 *Si un número natural puede expresarse como suma de tres cuadrados de números racionales, entonces es suma de tres cuadrados de números enteros.*

[1] En realidad es posible demostrar analíticamente un resultado de este tipo que permite completar la prueba.

DEMOSTRACIÓN: Si n es suma de tres cuadrados racionales, entonces

$$n = \frac{x_1^2}{w^2} + \frac{x_2^2}{w^2} + \frac{x_3^2}{w^2},$$

donde x_1, x_2, x_3, w son números naturales y $w > 0$. Equivalentemente,

$$x_1^2 + x_2^2 + x_3^2 = nw^2.$$

Podemos considerar la representación de este tipo que tiene w mínimo, y vamos a probar que tiene que ser $w = 1$, con lo que n será suma de tres cuadrados enteros.

Si w divide a todos los x_i, dividiendo la ecuación entre w^2 obtenemos una representación de n como suma de tres cuadrados enteros. Supongamos, pues, que existe un i tal que $w \nmid x_i$. Sea y_i el entero más próximo a x_i/w, de modo que $|y_i - x_i/w| \le 1/2$ o, equivalentemente,

$$|wy_i - x_i| \le \frac{w}{2}.$$

Como w no divide a un x_i, para dicho índice se cumple que $y_i w \ne x_i$. Llamemos

$$a = n - y_1^2 - y_2^2 - y_3^2, \qquad b = 2nw - 2x_1y_1 - 2x_2y_2 - 2x_3y_3$$

y sea $z_i = ax_i + by_i$. Entonces $z_i^2 = a^2x_i^2 + b^2y_i^2 + 2abx_iy_i$, luego

$$z_1^2 + z_2^2 + z_3^2 = a^2(x_1^2 + x_2^2 + x_3^2) + b^2(y_1^2 + y_2^2 + y_3^2) + 2ab(x_1y_1 + x_2y_2 + x_3y_3)$$

$$= a^2nw^2 + b^2(a+n) + ab(2nw - b) = n(aw+b)^2.$$

Basta probar que $0 \le aw + b < w$, pues esto contradice la minimalidad de w. Para ello observamos que

$$(wy_1 - x_1)^2 + (wy_2 - x_2)^2 + (wy_3 - x_3)^2 \le \frac{3w^2}{4}$$

y, por otra parte,

$$\frac{3w^2}{4} \ge w^2(y_1^2 + y_2^2 + y_3^2) + x_1^2 + x_2^2 + x_3^2 - 2w(x_1y_1 + x_2y_2 + x_3y_3)$$

$$= w^2(a+n) + nw^2 + w(b - 2nw) = w(aw+b) \ge 0.$$

(La última desigualdad se debe a que $w(aw+b)$ es una suma de cuadrados.) De aquí se sigue que, en efecto, $0 \le aw + b \le 3w/4 < w$. ∎

Ahora ya podemos probar:

Teorema 7.8 *Un número natural es suma de tres cuadrados si y sólo si no es de la forma* $4^k(8l+7)$.

DEMOSTRACIÓN: Veamos que basta probar que si m es libre de cuadrados y $m \not\equiv 7 \pmod 8$, entonces m es suma de tres cuadrados. En efecto, admitiendo esto, consideremos un número n que no sea de la forma indicada en el enunciado. Podemos expresarlo en la forma $n = 2^{2k}a^2m$, donde a es impar y m es libre de cuadrados. Entonces se cumple que $m \not\equiv 7 \pmod 8$, pues en caso contrario, como $a^2 \equiv 1 \pmod 8$, sería $a^2m \equiv 7 \pmod 8$ y n tendría la forma del enunciado. Por el caso particular que suponemos, resulta que m es suma de tres cuadrados, luego n también lo es.

Suponemos, pues, que m es libre de cuadrados y que $m \not\equiv 7 \pmod 8$. Basta probar que existe un número natural $r > 0$, que cumpla:

1. r es libre de cuadrados,
2. $(r, m) = 1$,
3. r es suma de dos cuadrados,
4. m es un resto cuadrático módulo r,
5. $-r$ es un resto cuadrático módulo m.

En efecto, si existe tal r, por el teorema de Legendre 3.11, la ecuación

$$x^2 + ry^2 - mz^2 = 0$$

tiene una solución entera no trivial. Si fuera $z = 0$, entonces también $x = y = 0$, luego $z \neq 0$ y podemos dividir:

$$m = (x/z)^2 + r(y/z)^2,$$

pero r es suma de dos cuadrados: $r = r_1^2 + r_2^2$, luego

$$m = (x/z)^2 + (r_1y/z)^2 + (r_2y/z)^2.$$

Por el teorema anterior, m es suma de tres cuadrados.

Vamos a construir r. Para ello pongamos que $m = 2^e m_1$, donde $e = 0, 1$ y $m_1 = p_1 \cdots p_k$ es un producto de primos impares distintos. Definimos

$$f = \begin{cases} 0 & \text{si } e = 1 \text{ o bien } e = 0,\ m_1 \equiv 1, 5 \pmod 8, \\ 1 & \text{si } e = 0 \text{ y } m_1 \equiv 3 \pmod 8. \end{cases}$$

Por el teorema chino del resto existe un entero q tal que

$$\left(\frac{q}{p_i}\right) = \left(\frac{-2^f}{p_i}\right), \qquad i = 1, \ldots, k$$

y

$$q \equiv \begin{cases} 1 \pmod 8 & \text{si } m_1 \equiv 1, 5 \pmod 8, \\ 5 \pmod 8 & \text{si } m_1 \equiv 3 \pmod 8. \end{cases}$$

(Basta tomar q congruente módulo cada p_i con un resto cuadrático o no cuadrático según el valor del símbolo de Legendre de la derecha y congruente módulo 8 con el valor requerido.) Sabemos que q está determinado salvo múltiplos de m_1 y claramente $(q, 8m_1) = 1$, luego el teorema de Dirichlet nos permite sustituir q por otro $q + 8m_1t$ que sea primo, es decir, que podemos suponer que q es primo.

Llamamos $r = 2^f q > 0$. Notemos que $(r, m) = 1$, pues si $2 \mid m$ entonces $f = 0$. Además, como $q \equiv 1 \pmod 4$, es claro que r es suma de dos cuadrados. Falta probar que cumple las dos últimas propiedades:

$$\left(\frac{m}{q}\right) = \left(\frac{2^e}{q}\right)\prod_{i=1}^{k}\left(\frac{p_i}{q}\right) = \left(\frac{2^e}{q}\right)\prod_{i=1}^{k}\left(\frac{q}{p_i}\right) = \left(\frac{2^e}{q}\right)\prod_{i=1}^{k}\left(\frac{-2^f}{p_i}\right)$$

$$= \left(\frac{2^e}{q}\right)\left(\frac{-2^f}{m_1}\right) = 1.$$

Veamos la última igualdad. Si $e = 1$, entonces $f = 0$, con lo que tenemos

$$\left(\frac{2}{q}\right)\left(\frac{-1}{m_1}\right),$$

y, por la elección de q, ambos factores valen 1 si $m_1 \equiv 1 \pmod 4$ y ambos valen -1 si $m_1 \equiv -1 \pmod 4$. Supongamos ahora que $e = 0$, y entonces tenemos[2]

$$\left(\frac{-1}{m_1}\right)\left(\frac{2^f}{m_1}\right),$$

y de nuevo ambos factores son iguales por la definición de f. Esto prueba que m es un cuadrado módulo q y, como también lo es módulo 2 si es impar, concluimos que m es un cuadrado módulo r.

Para probar que $-r$ es un cuadrado módulo m basta probar que lo es módulo cada p_i (trivialmente lo es módulo 2). En efecto:

$$\left(\frac{-r}{p_i}\right) = \left(\frac{-2^f}{p_i}\right)\left(\frac{q}{p_i}\right) = \left(\frac{-2^f}{p_i}\right)\left(\frac{-2^f}{p_i}\right) = 1.$$

■

El teorema anterior lo demostró Legendre en 1798 (sin usar el teorema de Dirichlet ni la ley de reciprocidad). Un poco antes, en 1796, con 19 años, Gauss había demostrado el teorema siguiente, que se conoce como el *teorema Eureka,* porque Gauss anotó en su diario: "$E\Upsilon PHKA!$ Num $= \Delta + \Delta + \Delta$:

Teorema 7.9 *Todo número natural es suma de tres números triangulares.*

[2]Notemos que si no hubiéramos exceptuado el caso $m \equiv 7 \pmod 8$ tendríamos que $(-1/m_1) = -1$, con lo que tendría que ser $f = 1$, pero igualmente $(2/m_1) = 1$ y m no sería un cuadrado módulo r.

DEMOSTRACIÓN: Dado un número natural n, queremos probar que existen números naturales k, r, s tales que

$$n = \frac{k(k+1)}{2} + \frac{r(r+1)}{2} + \frac{s(s+1)}{2}.$$

Desarrollando, esto equivale a que

$$2n = k^2 + k + r^2 + r + s^2 + s$$

o también:

$$8n = 4k^2 + 4k + 4r^2 + 4r + 4s^2 + 4s = (2k+1)^2 + (2r+1)^2 + (2s+1)^2 - 3.$$

Ahora bien, por el teorema anterior tenemos que existen enteros x, y, z tales que

$$8n + 3 = x^2 + y^2 + z^2$$

y, como $x^2 + y^2 + z^2 \equiv 3 \pmod 8$, los tres enteros tienen que ser impares. Haciendo $x = 2k+1$, $y = 2r+1$, $z = 2s+1$ tenemos que k, r, s cumplen lo requerido. ∎

Nota Observemos que el teorema 7.8 implica el teorema 3.5 según el cual todo número natural es suma de cuatro cuadrados, si bien este teorema (que fue demostrado por Lagrange en 1770) tiene una prueba mucho más simple y directa.

En efecto, basta probar que todo número de la forma $4^k(8l+7)$ es suma de cuatro cuadrados, pero para ello basta ver que lo es todo número de la forma $8l+7$, y ello se debe a que $8l + 7 = (8l+3) + 2^2$, y el primer sumando es suma de tres cuadrados. ∎

7.8 El signo de las sumas de Gauss

La fórmula (7.1) es válida para cualquier suma de Gauss $G(p)$ construida a partir de cualquier raíz p-ésima de la unidad $\zeta \neq 1$ en cualquier cuerpo. Si tomamos, más concretamente, $\zeta \in \mathbb{C}$, de ella se desprende que

$$G(p) = \begin{cases} \pm\sqrt{p} & \text{si } p \equiv 1 \pmod 4, \\ \pm\sqrt{p}\,i & \text{si } p \equiv -1 \pmod 4. \end{cases}$$

El signo depende de la elección de ζ, pero Gauss constató en ejemplos concretos que, cuando tomamos precisamente $\zeta = e^{2\pi i/p}$, el signo siempre es positivo, si bien le costó más de cuatro años encontrar una demostración general. Hoy en día se conocen varias pruebas más o menos conceptuales basadas en el álgebra lineal, en la teoría de funciones holomorfas, etc. Aquí vamos a ver una prueba elemental debida a Cauchy:

Teorema 7.10 *Si p es un primo impar y $\zeta = e^{2\pi i/p}$, entonces*

$$G(p) = \sum_{k=1}^{p-1} \left(\frac{k}{p}\right) \zeta^k = \begin{cases} \sqrt{p} & si\ p \equiv\ \ 1\ (\text{mód } 4), \\ \sqrt{p}\, i & si\ p \equiv -1\ (\text{mód } 4). \end{cases}$$

Demostración: La clave de la prueba es considerar el número

$$H(p) = \prod_{k=1}^{(p-1)/2} (\zeta^{2k-1} - \zeta^{-(2k-1)}).$$

Por ejemplo, para $p = 5$ tenemos

$$H(5) = (\zeta - \zeta^{-1})(\zeta^3 - \zeta^{-3}) = \zeta - \zeta^2 - \zeta^3 + \zeta^4$$
$$= \left(\frac{1}{5}\right)\zeta + \left(\frac{2}{5}\right)\zeta^2 + \left(\frac{3}{5}\right)\zeta^3 + \left(\frac{4}{5}\right)\zeta^4 = G(5).$$

Esto no es casual. Vamos a probar que $H(p) = G(p)$, con la diferencia de que en el caso de $H(p)$ es fácil calcular el signo que en el caso de $G(p)$ se resiste a ser calculado. La parte más fácil es observar que

$$(\zeta^{2k-1} - \zeta^{-(2k-1)})^2 = -(\zeta^{-(2k-1)} - \zeta^{2k-1})(\zeta^{2k-1} - \zeta^{-(2k-1)})$$
$$= -\zeta^{-(2k-1)}(1 - \zeta^{4k-2})\zeta^{2k-1}(1 - \zeta^{-(4k-2)}) = -(1 - \zeta^{4k-2})(1 - \zeta^{-(4k-2)}),$$

luego

$$H(p)^2 = (-1)^{(p-1)/2} \prod_{k=1}^{(p-1)/2} (1 - \zeta^{4k-2})(1 - \zeta^{-(4k-2)}) = (-1)^{(p-1)/2} \prod_{k=1}^{p-1} (1 - \zeta^k),$$

pues, en el penúltimo término, el primer factor recorre todas las potencias pares de ζ, luego el segundo recorre todas las potencias impares, así que en total están todas. Ahora basta tener en cuenta que, por la igualdad

$$x^p - 1 = (x - 1)(x^{p-1} + x^{p-2} + \cdots + x + 1),$$

todas las potencias ζ^k son raíces del polinomio

$$c_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1 = \prod_{k=1}^{p-1} (x - \zeta^k),$$

y evaluando en 1 queda que $H(p)^2 = (-1)^{(p-1)/2} p$.

Teniendo en cuenta (7.1), tenemos que $H(p)^2 = G(p)^2$, luego $G(p) = \pm H(p)$.

En segundo lugar observamos que $\zeta^{-(2k-1)}$ es el conjugado de ζ^{2k-1} (pues son dos números complejos mutuamente inversos de módulo 1). Por lo tanto,

$$\zeta^{2k-1} - \zeta^{-(2k-1)} = 2i \operatorname{Im} \zeta^{2k-1} = 2i \operatorname{sen} \frac{2(2k-1)\pi}{p},$$

luego

$$H(p) = i^{(p-1)/2} \prod_{k=1}^{(p-1)/2} (2 \operatorname{sen} \frac{2(2k-1)\pi}{p}).$$

Veamos cuántos factores son negativos en el producto. El término $2k-1$ recorre todos los números impares entre 1 y $p-2$, luego

$$0 < \frac{2\pi}{p} \le \frac{2(2k-1)\pi}{p} \le \frac{2(p-2)\pi}{p} < 2\pi,$$

luego el seno será negativo cuando

$$\pi < \frac{2(2k-1)\pi}{p} < 2\pi,$$

lo que equivale a que

$$\frac{p+2}{4} < k \le \frac{p-1}{2},$$

luego el número de factores negativos es $(p-1)/2 - E[(p+2)/4]$.

Si $p \equiv 1 \pmod 4$, digamos $p = 4r+1$, entonces

$$\frac{p-1}{2} - E[(p+2)/4] = 2r - E[r+3/4] = 2r - r = r$$

y, por otra parte, $i^{(p-1)/2} = (-1)^r$, de donde se sigue que $H(p) > 0$.

Si $p \equiv -1 \pmod 4$, digamos $p = 4r-1$, entonces

$$\frac{p-1}{2} - E[(p+2)/4] = 2r - 1 - E[r+1/4] = 2r - 1 - r = r - 1,$$

mientras que $i^{(p-1)/2} = i^{2r-1} = -i(-1)^r = i(-1)^{r-1}$, y en este caso concluimos que $H(p)/i > 0$.

Teniendo en cuenta que $H(p)^2 = (-1)^{(p-1)/2}p$, con esto hemos probado que

$$H(p) = \begin{cases} \sqrt{p} & \text{si } p \equiv \ \ 1 \pmod 4, \\ \sqrt{p}i & \text{si } p \equiv -1 \pmod 4. \end{cases}$$

Finalmente vamos a probar que $G(p) = H(p)$ o, equivalentemente, que en la relación $G(p) = \pm H(p)$ que hemos demostrado, el signo correcto es siempre el positivo.

Tenemos que ζ es raíz del polinomio

$$c_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1,$$

que hemos probado que es irreducible en 3.34, pero esto implica que $c_p(x)$ divide a todo polinomio $f(x) \in \mathbb{Q}[x]$ que tenga a ζ por raíz.[3]

[3]Esto es un hecho algebraico general que más adelante estudiaremos con detalle y que no deberíamos considerar parte de esta demostración. En efecto $c_p(x) \mid f(x)$, pues en caso contrario sería $(c_p(x), f(x)) = 1$ y, por la relación de Bezout, existirían polinomios tales que $u(x)c_p(x) + v(c)f(x) = 1$, y evaluando en ζ quedaría $0 = 1$.

Más aún, si $f(x) \in \mathbb{Z}[x]$, entonces el cociente también está en $\mathbb{Z}[x]$, porque, como $c_p(x)$ tiene coeficiente director 1, podemos dividir $f(x) = c_p(x)q(x) + r(x)$, donde todos los polinomios están en $\mathbb{Z}[x]$ y el resto tiene grado menor que $c_p(x)$, pero evaluando en ζ queda que $r(\zeta) = 0$, luego, según hemos probado, $c_p(x)$ divide a $r(x)$ en $\mathbb{Q}[x]$, lo cual exige que $r(x)$ sea nulo, y así $c_p(x)$ divide a $f(x)$ en $\mathbb{Z}[x]$.

Consideremos los polinomios

$$G(x) = \sum_{k=1}^{p-1} \left(\frac{k}{p}\right) x^k, \qquad H(x) = \prod_{k=1}^{(p-1)/2} (x^{2k-1} - x^{p-(2k-1)}),$$

Hemos probado que $G(\zeta) \mp H(\zeta) = 0$, luego existe un polinomio $d(x) \in \mathbb{Z}[x]$ tal que

$$c_p(x)d(x) = G(x) \mp H(x).$$

Como todos los polinomios tienen coeficientes enteros, podemos considerar los polinomios de $\mathbb{Z}_p[x]$ que se obtienen al sustituir cada coeficiente por su clase de restos módulo p. Tenemos entonces que

$$\bar{c}_p(x)\bar{d}(x) = \bar{G}(x) \mp \bar{H}(x).$$

Pero en $\mathbb{Z}_p[x]$, la relación

$$(x - \bar{1})\bar{c}_p(x) = x^p - \bar{1} = (x - \bar{1})^p$$

se traduce en que $\bar{c}_p(x) = (x - \bar{1})^{p-1}$, por lo que

$$(x - \bar{1})^{p-1}\bar{d}(x) = \bar{G}(x) \mp \bar{H}(x).$$

Ahora consideramos el polinomio $\pi = x - \bar{1} \in \mathbb{Z}_p[x]$ con lo que la igualdad anterior equivale a

$$\pi^{p-1}\bar{d}(\pi + \bar{1}) = \bar{G}(\pi + \bar{1}) \mp \bar{H}(\pi + \bar{1}),$$

o también a que

$$\bar{G}(\pi + \bar{1}) \equiv \pm\bar{H}(\pi + \bar{1}) \text{ (mód } \pi^{p-1}). \tag{7.3}$$

Nuestro objetivo es demostrar que el signo tiene que ser positivo.

Examinamos el miembro derecho:

$$\bar{H}(\pi + \bar{1}) = \prod_{k=1}^{(p-1)/2} ((\pi + \bar{1})^{2k-1} - (\pi + \bar{1})^{p-(2k-1)}),$$

donde

$$(\pi + \bar{1})^{2k-1} - (\pi + \bar{1})^{p-(2k-1)} \equiv \bar{1} + \overline{2k-1}\pi - \bar{1} + \overline{2k-1}\pi = \overline{4k-2}\,\pi \text{ (mód } \pi^2).$$

(Hemos aplicado la fórmula del binomio de Newton truncando los términos en los que aparecen potencias de π con exponente mayor que 1). Por lo tanto

$$\begin{aligned} \bar{H}(\pi + \bar{1}) &= \prod_{k=1}^{(p-1)/2} \overline{4k-2}\,\pi + \pi^2 R_k(x)) \\ &\equiv \bar{2}^{(p-1)/2}(\bar{1} \cdot \bar{3} \cdots \overline{p-2})\pi^{(p-1)/2} \text{ (mód } \pi^{(p+1)/2}). \end{aligned} \tag{7.4}$$

(Al desarrollar el producto, todos los términos son divisibles entre $\pi^{(p+1)/2}$ menos el que hemos dejado explícito.)

Ahora examinamos el miembro izquierdo de (7.3):

$$\begin{aligned}\bar{G}(\pi+\bar{1}) &= \sum_{k=1}^{p-1}\left(\frac{k}{p}\right)(\pi+1)^k = \sum_{k=1}^{p-1}\left(\frac{k}{p}\right)\sum_{r=0}^{k}\binom{k}{r}\pi^r \\ &= \sum_{r=1}^{p-1}\sum_{k=r}^{p-1}\left(\frac{k}{p}\right)\binom{k}{r}\pi^r = \sum_{r=1}^{p-1} c_r\pi^r, \end{aligned} \tag{7.5}$$

donde

$$c_r = \frac{1}{r!}\sum_{k=r}^{p-1}\left(\frac{k}{p}\right)\bar{k}(\bar{k}-\bar{1})\cdots(\bar{k}-\bar{r}+\bar{1}) = \frac{1}{r!}\sum_{k=1}^{p-1}\left(\frac{k}{p}\right)\bar{k}(\bar{k}-\bar{1})\cdots(\bar{k}-\bar{r}+\bar{1}),$$

donde hay que entender que $1/r!$ es la clase inversa de $\overline{r!}$ en $\mathbb{Z}_p$ (notemos que, como $r < p$, no es la clase nula), y donde hemos extendido el sumatorio con sumandos nulos, pues todos los sumandos añadidos anulan al último polinomio. Multiplicando los r factores obtenemos una expresión de la forma

$$\begin{aligned} c_r &= \frac{1}{r!}\sum_{k=1}^{p-1}\left(\frac{k}{p}\right)(\bar{a}_r\bar{k}^r + \bar{a}_{r-1}\bar{k}^{r-1} + \cdots + \bar{a}_1\bar{k}) \\ &= \frac{1}{r!}\sum_{d=1}^{r}\bar{a}_d\sum_{k=1}^{p-1}\left(\frac{k}{p}\right)\bar{k}^d, \end{aligned} \tag{7.6}$$

donde $\bar{a}_r = 1$.

Vamos a probar que $c_r = 0$, para $r = 0, \ldots, (p-3)/2$. Para ello empezamos probando que

$$\sum_{k=1}^{p-1} k^d \equiv \begin{cases} 0 \pmod p & \text{si } 0 \le d < p-1, \\ -1 \pmod p & \text{si } d = p-1. \end{cases}$$

El caso $d = p-1$ es inmediato por el teorema de Fermat, según el cual $k^{p-1} \equiv 1 \pmod p$, luego la suma es congruente con $p-1$, luego con -1.

Para el caso restante tomamos una raíz primitiva g módulo p, de modo que $g^d \not\equiv 1 \pmod p$, y observamos que

$$\overline{g^d\sum_{k=1}^{p-1} k^d} = \sum_{k\in U_p}(\bar{g}k)^d = \sum_{k\in U_p} k^d,$$

pues cuando k recorre U_p, lo mismo le sucede a $\bar{g}k$, y así

$$(\bar{g}^d - \bar{1})\overline{\sum_{k=1}^{p-1} k^d} = 0,$$

y como $\bar{g}^d - \bar{1} \neq \bar{0}$, tiene que ser $\bar{0}$ la clase de la suma, luego ésta es divisible entre p.

A su vez, de aquí deducimos que

$$\sum_{k=1}^{p-1}\left(\frac{k}{p}\right)k^d \equiv \begin{cases} 0 \pmod p & \text{si } 0 \le d < (p-1)/2,\\ -1 \pmod p & \text{si } d = (p-1)/2.\end{cases}$$

Basta aplicar el criterio de Euler 5.8:

$$\sum_{k=1}^{p-1}\left(\frac{k}{p}\right)k^d \equiv \sum_{k=1}^{p-1} k^{(p-1)/2+d} \pmod p$$

y aplicar el resultado precedente, pues en el caso $d < (p-1)/2$ tenemos que $(p-1)/2 + d < p-1$, mientras que si $d = (p-1)/2$ el exponente es $p-1$.

Con esto ya podemos analizar la expresión (7.6), y vemos que, en efecto, si $r < (p-1)/2$, todos los sumatorios finales son nulos, mientras que para $r = (p-1)/2$ son todos nulos menos el correspondiente a $d = r$. En resumen: $c_r = \bar{0}$ para $1 \le r < (p-1)/2$ y

$$c_{(p-1)/2} = -\frac{1}{((p-1)/2)!}.$$

Por consiguiente, la igualdad (7.5) nos da ahora que

$$\bar{G}(\pi+1) \equiv -\frac{\pi^{(p-1)/2}}{((p-1)/2)!} \pmod{\pi^{(p+1)/2}}. \tag{7.7}$$

Por último, combinamos esta última congruencia con (7.3) y (7.4), con lo que resulta que

$$-\pi^{(p-1)/2} \equiv (\frac{p-1}{2})!\,\bar{2}^{(p-1)/2}(\bar{1}\cdot\bar{3}\cdots\overline{p-2})\pi^{(p-1)/2}\,\pi^{(p-1)/2} \pmod{\pi^{(p+1)/2}},$$

pero $((p-1)/2)!\cdot 2^{(p-1)/2} = 2\cdot 4\cdots(p-1)$, luego en total tenemos:

$$-\pi^{(p-1)/2} \equiv \overline{(p-1)!}\pi^{(p-1)/2}\,\pi^{(p-1)/2} \pmod{\pi^{(p+1)/2}},$$

Aplicamos el teorema de Wilson 3.10:

$$-\pi^{(p-1)/2} \equiv \pm(-1)\pi^{(p-1)/2} \pmod{\pi^{(p+1)/2}},$$

luego

$$\pi^{(p-1)/2} \equiv \pm\pi^{(p-1)/2} \pmod{\pi^{(p+1)/2}}.$$

Y ahora ya podemos concluir que el signo negativo es imposible, pues supondría que

$$\pi^{(p+1)/2} \mid \bar{2}\pi^{(p-1)/2},$$

cuando el polinomio de la izquierda tiene grado mayor que el de la derecha. ∎

En términos de la expresión (7.2), el teorema anterior admite un enunciado más elemental, sin el símbolo de Legendre:

Teorema 7.11 *Si p es un primo impar, entonces*

$$G(p) = \sum_{k=1}^{p} e^{2k^2\pi i/p} = \begin{cases} \sqrt{p} & \text{si } p \equiv 1 \pmod 4,\\ \sqrt{p}\,i & \text{si } p \equiv -1 \pmod 4.\end{cases}$$

Capítulo VIII

Números y enteros algebraicos

En los capítulos anteriores hemos demostrado propiedades diversas de los números enteros utilizando la aritmética de anillos como $\mathbb{Z}[i]$, $\mathbb{Z}\big[\sqrt{-2}\,\big]$, $\mathbb{Z}\big[\sqrt{-3}\,\big]$ o $\mathbb{Z}[\zeta]$. En el capítulo siguiente presentaremos una teoría general sobre este tipo de anillos, pero antes conviene que introduzcamos algunos conceptos básicos aún más generales para entenderla desde la perspectiva adecuada.

8.1 Los números complejos

En la sección 1.1 de [ITAn] definimos los números complejos como los pares (x, y) de números reales, y vimos que, con una notación adecuada, todo número complejo se expresa de forma única como $z = x + yi$, donde x, y son números reales, y que, con la suma y el producto definidos adecuadamente, los números complejos resultan ser un cuerpo.

Todo esto supone una serie de comprobaciones sencillas, pero laboriosas (hay que probar que la suma y el producto cumplen todas las propiedades de la definición de cuerpo, etc.). En realidad todo esto puede sustituirse por una simple aplicación del teorema 6.1 o, equivalentemente, calcando la definición 4.2 del cuerpo $\mathbb{Q}(i)$, pero sustituyendo $\mathbb{Q}$ por $\mathbb{R}$.

Explícitamente, podemos definir el cuerpo $\mathbb{C}$ de los números complejos como el cuerpo de clases de restos del anillo de polinomio $\mathbb{R}[x]$ módulo el polinomio irreducible $x^2 + 1$ y llamar $i = \bar{x}$. Así el teorema 6.1 nos da directamente, sin necesidad de cálculo alguno que:

1. $\mathbb{C}$ es un cuerpo.
2. $\mathbb{R}$ se identifica con un subcuerpo de $\mathbb{C}$.
3. Si llamamos $i = \bar{x} \in \mathbb{C}$, se cumple que $i^2 = -1$.
4. Todo número complejo se expresa de forma única como $z = x + yi$, donde $x, y \in \mathbb{R}$.

5. Como $-i$ es también raíz del polinomio x^2+1, la conjugación $\mathbb{C} \longrightarrow \mathbb{C}$ dada por $\overline{x+yi} = x - iy$ cumple

$$\overline{z_1+z_2} = \bar{z}_1 + \bar{z}_2, \qquad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2.$$

De las propiedades 1, 2, 3 se sigue a su vez trivialmente que la suma y el producto de números complejos vienen dados por

$$(x_1+y_1i)+(x_2+y_2i) = (x_1+x_2)+(y_1+y_2)i,$$

$$(x_1+y_1i)(x_2+y_2i) = (x_1x_2-y_1y_2)+(x_1y_2+x_2y_2)i.$$

Todos estos hechos son los que en [ITAn] reducíamos a comprobaciones laboriosas, y que con este "planteamiento algebraico" resultan triviales. Y estos mismos hechos vuelven evidente que el cuerpo $\mathbb{C}$ construido a partir del teorema 6.1 es esencialmente el mismo construido en [ITAn], en cuanto a que, aunque no estamos llamando "números complejos" exactamente a los mismos objetos (en un caso son pares de números reales y en otro clases de restos de polinomios), lo cierto es que en ambos casos los números complejos terminan siendo expresiones de la forma $x+yi$ que se suman y se multiplican con los mismos criterios.

Pero en [ITAn] demostramos una propiedad fundamental de los números complejos que requiere necesariamente técnicas analíticas en su demostración. Se trata del teorema [ITAn 3.27]:

Teorema 8.1 (Teorema fundamental del álgebra) *Todo polinomio no constante con coeficientes en $\mathbb{C}$ tiene al menos una raíz en $\mathbb{C}$.*

Desde un punto de vista algebraico, esto significa que los polinomios irreducibles del anillo $\mathbb{C}[x]$ son exactamente los polinomios de grado 1, pues un polinomio $p(x)$ de grado mayor que 1 que tenga una raíz α en $\mathbb{C}$ no puede ser irreducible en $\mathbb{C}[x]$, ya que se descompone como $p(x) = (x-\alpha)q(x)$, donde $q(x)$ es otro polinomio de grado no nulo, luego no unitario.

Por lo tanto, la descomposición en factores irreducibles de cualquier polinomio de $\mathbb{C}[x]$ es de la forma

$$p(x) = a(x-a_1)\cdots(x-a_n),$$

donde a es, concretamente, el coeficiente director de $p(x)$

Una consecuencia del teorema fundamental es que todos los cuerpos que podemos obtener a partir del teorema 6.1 partiendo del cuerpo $\mathbb{Q}$ de los números racionales pueden identificarse con subcuerpos de $\mathbb{C}$.

En efecto, si $p(x)$ es un polinomio irreducible en $\mathbb{Q}[x]$, el teorema 6.1 nos da un cuerpo K que contiene a $\mathbb{Q}$ en el que $p(x)$ tiene una raíz α, pero por otra parte el teorema fundamental nos dice que $p(x)$ también tiene una raíz α' en $\mathbb{C}$, y el teorema 6.1 nos dice entonces que existe una única aplicación $f : K \longrightarrow \mathbb{C}$ que nos permite identificar cada elemento de K con un número complejo.

Por ejemplo, el cuerpo $\mathbb{Q}(i)$ construido en 4.2 puede identificarse con el cuerpo de los números complejos de la forma $a+bi$, donde a, b son números racionales y, en particular, el anillo $\mathbb{Z}[i]$ de los enteros de Gauss se puede identificar con el anillo de los números complejos de la forma $a+bi$, donde a, b son números enteros.

Similarmente, el polinomio x^2+x+1 tiene raíces en $\mathbb{C}$. Una de ellas es

$$\zeta = \frac{-1+\sqrt{3}i}{2}$$

y entonces podemos identificar el cuerpo $\mathbb{Q}(\zeta)$ definido en 6.2 con el cuerpo formado por los números complejos de la forma $a+b\zeta$, donde a, b son números racionales y, en particular, el anillo $\mathbb{Z}[\zeta]$ de los enteros de Eisenstein se puede identificar con el anillo de todos los números complejos de la forma $a+b\zeta$, donde a, b son números enteros.

En principio, considerar que $\mathbb{Z}[\zeta]$ es el "anillo abstracto" construido en 6.2 o bien que está formado concretamente por números complejos no tiene relevancia alguna si vamos a trabajar exclusivamente con enteros de Eisenstein, como hemos hecho en el capítulo VI, pero considerar que todos los números que estamos considerando son realmente números complejos tiene la ventaja de que ahora podemos operarlos entre sí.

Por ejemplo, hasta ahora $\mathbb{Q}(i)$ y $\mathbb{Q}(\zeta)$ eran dos cuerpos sin relación alguna entre sí, de modo que no tenía ningún sentido plantearse la suma $i+\zeta$, mientras que esta suma tiene perfecto sentido si entendemos que i y ζ son dos números complejos. Concretamente, con la elección de ζ que hemos hecho anteriormente, resulta que

$$i+\zeta = i + \frac{-1+\sqrt{3}i}{2} = -\frac{1}{2} + \frac{2+\sqrt{3}}{2}i,$$

y este número complejo no es ni un entero de Gauss ni un entero de Eisenstein, pero es un número complejo perfectamente definido.

8.2 Polinomios simétricos

En la sección siguiente probaremos los resultados básicos sobre los números algebraicos, y para ello nos basaremos en algunos hechos sobre polinomios simétricos que presentamos aquí:

Las fórmulas de Vieta Observemos que:

$$\begin{aligned}
(x-a)(x-b) &= x^2-(a+b)x+ab\\
(x-a)(x-b)(x-c) &= x^3-(a+b+c)x^2+(ab+ac+bc)x-abc\\
(x-a)(x-b)(x-c)(x-d) &= x^4-(a+b+c+d)x^3\\
&+ (ab+ac+ad+bc+bd+cd)x^2\\
&- (abc+abd+acd+bcd)x+abcd.
\end{aligned}$$

Es fácil captar el patrón general que siguen estas fórmulas, y que expresan los coeficientes de un polinomio en términos de sus raíces (suponiendo que tiene tantas como indica su grado, contándolas con sus multiplicidades).

Definición 8.2 Si A es un dominio, llamaremos *polinomios simétricos elementales* de $A[x_1,\dots,x_n]$ a los polinomios $e_0,\dots,e_n$ dados por

$$e_0 = 1, \quad e_k = \sum_{1\le i_1<\cdots<i_k\le n} x_{i_1}\cdots x_{i_k} \qquad \text{para } k = 1,\dots,n.$$

Por ejemplo, los polinomios simétricos de 3 indeterminadas son

$$1, \qquad x+y+z, \qquad xy+xz+yz, \qquad xyz.$$

En otras palabras, el polinomio e_k es la suma de todos los monomios que pueden construirse multiplicando k variables distintas.

En general, un polinomio es *simétrico* si permanece inalterado cuando se intercambian sus indeterminadas. Es claro que los polinomios simétricos elementales son simétricos en este sentido. Es fácil probar:

Teorema 8.3 *Sea A un dominio, $n>1$, $e_0,\dots,e_n$ los polinomios simétricos elementales en $A[x_1,\dots,x_n]$ y $\bar e_0,\dots,\bar e_{n-1}$ los polinomios simétricos elementales en $A[x_1,\dots,x_{n-1}]$. Entonces:*

1. $e_n(x_1,\dots,x_n) = x_n\bar e_{n-1}(x_1,\dots,x_{n-1})$.
2. $e_k(x_1,\dots,x_n) = \bar e_k(x_1,\dots,x_{n-1}) + x_n\bar e_{k-1}(x_1,\dots,x_{n-1})$, *para* $1\le k<n$.

Ahora ya podemos probar las fórmulas de Vieta:

Teorema 8.4 (Viète) *Sea A un dominio y $e_0,\dots,e_n$ los polinomios simétricos elementales en $A[x_1,\dots,x_n]$. Entonces*

$$(x-x_1)\cdots(x-x_n) = \sum_{k=0}^{n}(-1)^{n-k}e_{n-k}(x_1,\dots,x_n)x^k.$$

Demostración: Por inducción sobre n. Para $n=1$ es obvio. Supongámoslo para $n-1$. Sean $\bar e_0,\dots,\bar e_{n-1}$ los polinomios simétricos elementales en $A[x_1,\dots,x_{n-1}]$. Entonces

$$\begin{aligned}
(x-x_1)\cdots(x-x_n) &= \sum_{k=0}^{n-1}(-1)^{n-1-k}\ \bar e_{n-1-k}x^k(x-x_n)\\
&= \sum_{k=0}^{n-1}(-1)^{n-1-k}\ \bar e_{n-1-k}x^{k+1} - \sum_{k=0}^{n-1}(-1)^{n-1-k}x_n\bar e_{n-1-k}x^k\\
&= (-1)^n x_n\bar e_{n-1} + x^n + \sum_{k=0}^{n-2}(-1)^{n-1-k}\ \bar e_{n-1-k}x^{k+1}\\
&\quad - \sum_{k=1}^{n-1}(-1)^{n-1-k}x_n\bar e_{n-1-k}x^k
\end{aligned}$$

$$\begin{aligned} &= (-1)^n x_n \bar{e}_{n-1} + x^n + \sum_{k=1}^{n-1} (-1)^{n-k} (\bar{e}_{n-k} + x_n \bar{e}_{n-1-k}) x^k \\ &= (-1)^n e_n + x^n + \sum_{k=1}^{n-1} (-1)^{n-k} e_{n-k} x^k = \sum_{k=0}^{n} (-1)^{n-k} e_{n-k} x^k. \end{aligned}$$

■

Para terminar vamos a demostrar que todos los polinomios simétricos pueden expresarse en términos de los polinomios simétricos elementales. Por ejemplo, $x^2 + y^2 + z^2$ es un polinomio simétrico, pero no es elemental. Sin embargo:

$$x^2 + y^2 + z^2 = (x + y + z)^2 - 2(xy + xz + yz) = e_1^2 - 2e_2.$$

En general:

Teorema 8.5 *Sea A un dominio y sean $e_1, \ldots, e_n$ los polinomios simétricos elementales en $A[x_1, \ldots, x_n]$. Cada polinomio simétrico $p(x_1, \ldots, x_n)$ con coeficientes en A se expresa de forma única como $g(e_1, \ldots, e_n)$, para un cierto polinomio $g(x_1, \ldots, x_n)$ con coeficientes en A.*

DEMOSTRACIÓN: Razonamos por inducción sobre n. Para $n = 1$ todo polinomio de $A[x_1]$ es simétrico y, como $e_1 = x$, el teorema se cumple trivialmente.

Supongamos el teorema para $n-1$. Sea $p(x_1, \ldots, x_n)$ un polinomio simétrico con coeficientes en A y ahora razonamos por inducción sobre el grado de p. En caso de que p sea de grado 0 la conclusión es trivial. Supongamos que todo polinomio simétrico de grado menor que m depende polinómicamente de $e_1, \ldots, e_n$ y que p tiene grado m.

Sea $\bar{p}(x_1, \ldots, x_{n-1}) = p(x_1, \ldots, x_{n-1}, 0) \in A[x_1, \ldots, x_{n-1}]$. Claramente $\bar{p}$ es simétrico, luego, por la primera hipótesis de inducción,

$$\bar{p} = g(\bar{e}_1(x_1, \ldots x_{n-1}), \ldots, \bar{e}_{n-1}(x_1, \ldots x_{n-1})),$$

para cierto polinomio g, donde $\bar{e}_1, \ldots, \bar{e}_{n-1}$ son los polinomios simétricos elementales de $A[x_1, \ldots, x_{n-1}]$.

Consideremos el polinomio $h(x_1, \ldots, x_n) = p(x_1, \ldots, x_n) - g(e_1, \ldots, e_{n-1})$, que también es simétrico. Por 8.3 se cumple que

$$e_i(x_1, \ldots, x_{n-1}, 0) = \bar{e}_i(x_1, \ldots, x_{n-1}) \qquad i = 1, \ldots, n-1,$$

luego

$$h(x_1, \ldots, x_{n-1}, 0) = \bar{p}(x_1, \ldots, x_{n-1}) - g(\bar{e}_1, \ldots, \bar{e}_{n-1}) = 0.$$

Viendo a h como polinomio en x_n con coeficientes en $A[x_1, \ldots, x_{n-1}]$, esto implica que x_n divide a $h(x_1, \ldots, x_n)$ y, por la simetría, todas las variables lo dividen también, luego su producto también, es decir, que e_n divide a h. Pongamos que $h = e_n \bar{h}(x_1, \ldots, x_n)$, donde claramente $\bar{h}$ también es simétrico y de grado menor que m, luego por la segunda hipótesis de inducción

$$\bar{h} = g^*(e_1(x_1, \ldots, x_n), \ldots, e_n(x_1, \ldots, x_n)),$$

para cierto polinomio g^*, con lo que $p = g(e_1, \ldots, e_{n-1}) + e_n g^*(e_1, \ldots, e_n)$ se expresa también polinómicamente en términos de $e_1, \ldots, e_n$.

Veamos ahora la unicidad. Hay que ver que si $g(e_1,\dots,e_n) = h(e_1,\dots,e_n)$, entonces $g = h$. Como $(g-h)(e_1,\dots,e_n) = 0$, llamando $f = g - h$, lo que tenemos que probar es que si $f(e_1,\dots,e_n) = 0$, entonces $f = 0$.

Nuevamente razonamos por inducción sobre n. Para $n = 1$ es trivial, pues $f(e_1) = 0$ es lo mismo que $f(x) = 0$ y, supuesto cierto para $n-1$, razonamos por inducción sobre el grado de f. Si f tiene grado 0 es obvio que tiene que ser $f = 0$. Suponemos que es cierto para polinomios de grado menor que el de f.

Haciendo $x_n = 0$ en la igualdad

$$f(e_1(x_1,\dots,x_n),\dots,e_n(x_1,\dots,x_n)) = 0$$

obtenemos que $f(\bar{e}_1,\dots,\bar{e}_{n-1},0) = 0$, donde $\bar{e}_i$ son los polinomios simétricos elementales en $n-1$ indeterminadas. Por la primera hipótesis de inducción concluimos que $f(x_1,\dots,x_{n-1},0) = 0$, luego x_n divide a f. Pongamos que $f = x_n g$, donde el grado de g es menor que el de f. Entonces

$$e_n(x_1,\dots,x_n)g(e_1(x_1,\dots,x_n),\dots,e_n(x_1,\dots,x_n)) = f(e_1,\dots,e_n) = 0,$$

luego $g(e_1,\dots,e_n) = 0$ y, por la segunda hipótesis de inducción, $g = 0$, luego también $f = 0$. ∎

8.3 Números algebraicos

Hemos visto que todos los cuerpos construidos a partir de $\mathbb{Q}$ mediante el teorema 6.1 pueden verse como subcuerpos de $\mathbb{C}$, pero sucede que no todos los números complejos pueden aparecer en cuerpos construidos de este modo. Para entender por qué, conviene introducir un nuevo concepto:

Definición 8.6 Se dice que un número complejo α es *algebraico* si existe un polinomio no nulo $p(x)$ con coeficientes racionales tal que $p(\alpha) = 0$. En caso contrario se dice que es *trascendente.*

Por ejemplo, todo número racional a es algebraico, ya que es raíz del polinomio $x - a$. También es algebraico i (porque es raíz del polinomio x^2+1) o $\sqrt{2}$ (es raíz de x^2-2) o ζ (es raíz de x^2+x+1), o $\sqrt[5]{7}$ (es raíz de x^5-7), etc.

Pero no todos los números complejos son algebraicos. Aunque no es trivial en absoluto, e y π son los ejemplos "más famosos" de números trascendentes.

Observemos que podríamos haber definido equivalentemente un número algebraico como un número complejo que es raíz de un polinomio no nulo con coeficientes enteros, pues si $p(x)$ es un polinomio con coeficientes racionales, multiplicándolo por el producto de los denominadores de sus coeficientes obtenemos un polinomio con coeficientes enteros y que tiene las mismas raíces.

Sin embargo, en la práctica es preferible otro arreglo alternativo, y es dividir $p(x)$ entre su coeficiente director para tener un polinomio mónico con coeficientes racionales con las mismas raíces. Considerar polinomios mónicos es la forma más conveniente de tener la unicidad en el teorema siguiente:

Teorema 8.7 *Si α es un número algebraico, existe un único polinomio mónico irreducible $p(x)$ en $\mathbb{Q}[x]$ que tiene a α por raíz. Si $f(x)$ es cualquier otro polinomio de $\mathbb{Q}[x]$ que tiene a α por raíz, entonces $p(x) \mid f(x)$.*

DEMOSTRACIÓN: Que α sea algebraico significa que es raíz de un polinomio $p(x)$ de $\mathbb{Q}[x]$ no nulo. Podemos tomarlo del menor grado posible.

Entonces $p(x)$ es irreducible, pues, ciertamente no es unitario (o no tendría raíces), y si pudiera descomponerse como $p(x) = u(x)v(x)$, entonces α sería raíz de uno de los factores, pero como $p(x)$ tiene grado mínimo, dicho factor tendría que tener el mismo grado que $p(x)$, luego el otro factor tiene que tener grado 0, luego es una unidad y $p(x)$ es irreducible.

Supongamos ahora que $f(x)$ es cualquier otro polinomio de $\mathbb{Q}[x]$ tal que $f(\alpha) = 0$, podemos dividir

$$f(x) = p(x)c(x) + r(x),$$

donde $r(x)$ es nulo o tiene grado menor que $p(x)$. Pero evaluando en α queda que $r(\alpha) = 0$, y como $p(x)$ tiene grado mínimo entre los polinomios no nulos que se anulan en α tiene que ser $r(x) = 0$, luego $p(x) \mid c(x)$.

Esto implica en particular la unicidad de $p(x)$, pues si $q(x)$ es cualquier polinomio mónico irreducible que tenga a α por raíz, entonces $p(x) \mid q(x)$, pero como $q(x)$ es irreducible esto implica que $q(x) = cp(x)$, para cierto $c \in \mathbb{Q}$ no nulo, pero como ambos polinomios son mónicos tiene que ser $c = 1$. ∎

Definición 8.8 Si α es un número algebraico, el polinomio dado por el teorema anterior se llama *polinomio mínimo* de α y se representa por $\operatorname{pol\,mín}\alpha$.

Por ejemplo,

$$\operatorname{pol\,mín} i = x^2 + 1, \qquad \operatorname{pol\,mín}\sqrt{-2} = x^2 + 2, \qquad \operatorname{pol\,mín}\zeta = x^2 + x + 1,$$

etc.

Esto nos permite definir el *grado* de un número algebraico como el grado de su polinomio mínimo.

Por ejemplo, los polinomios de grado 1 en $\mathbb{Q}[x]$ son todos irreducibles y, si son mónicos, son necesariamente de la forma $x - a$, con $a \in \mathbb{Q}$, de donde los números algebraicos de grado 1 son precisamente los números racionales.

En cambio, $i, \sqrt{-2}$ o ζ son ejemplos de números cuadráticos, es decir, de grado 2, como lo serán casi todos los números algebraicos que vamos a considerar en este libro, pero existen números algebraicos de todos los grados. El teorema 3.34 prueba al menos que hay números algebraicos de grado arbitrariamente alto.

Aquí vamos a probar un único hecho general sobre números algebraicos:

Teorema 8.9 *El conjunto $\mathbb{A}$ de todos los números complejos algebraicos es un cuerpo.*

Esto significa que al operar números algebraicos obtenemos nuevamente números algebraicos. Por ejemplo, es inmediato que los números reales $\sqrt{2}$ y $\sqrt{3}$ son algebraicos, luego $\sqrt{2}+\sqrt{3}$ también lo es, aunque sea menos inmediato encontrar un polinomio no nulo que se anule en la suma.

DEMOSTRACIÓN: Sean α y β dos números algebraicos y sea $p(x)$ el producto de sus polinomios mínimos, que es un polinomio mónico con coeficientes racionales que tiene a α y β por raíces (si α y β tienen el mismo polinomio mínimo, basta tomar como $p(x)$ dicho polinomio, sin necesidad de repetirlo). Como los polinomios irreducibles en $\mathbb{C}[x]$ tienen grado 1, la descomposición de $p(x)$ en factores irreducibles es de la forma

$$p(x) = (x-\alpha_1)\cdots(x-\alpha_n),$$

donde $\alpha = \alpha_i$ y $\beta = \alpha_j$ para ciertos índices i, j, no necesariamente distintos. El polinomio

$$r(x) = \prod_{ij}(x-\alpha_i-\alpha_j)$$

es mónico y tiene a $\alpha+\beta$ entre sus raíces. Si probamos que tiene coeficientes racionales, tendremos probado que $\alpha+\beta$ es algebraico. Para ello consideramos el polinomio

$$R(x, x_1, \ldots, x_n) = \prod_{ij}(x-x_i-x_j) \in \mathbb{Z}[x, x_1, \ldots, x_n],$$

que podemos considerar también como elemento del anillo $\mathbb{Z}[x][x_1, \ldots, x_n]$, es decir, como polinomio en las indeterminadas $x_1, \ldots x_n$ con coeficientes en el anillo $\mathbb{Z}[x]$. Claramente es simétrico en el sentido definido en 8.2, es decir, que si intercambiamos dos de sus indeterminadas x_i y x_j obtenemos el mismo polinomio (los factores se desordenan, pero el producto es el mismo). Por lo tanto, el teorema 8.5 nos asegura que

$$R(x, x_1, \ldots, x_n) = g(x, e_1, \ldots, e_n),$$

donde los polinomios $e_1, \ldots, e_n$ son los polinomios simétricos elementales en $x_1, \ldots, x_n$ y $g(x, t_1, \ldots, t_n)$ es un polinomio en $t_1, \ldots, t_n$ con coeficientes en el anillo $\mathbb{Z}[x]$. Ahora basta observar que, por 8.4, los números $a_i = e_i(\alpha_1, \ldots, \alpha_n)$ son, salvo el signo, los coeficientes de $p(x)$, por lo que son números racionales. Por consiguiente,

$$r(x) = R(x, \alpha_1, \ldots, \alpha_n) = g(x, a_1, \ldots, a_n)$$

tiene coeficientes racionales.

El mismo argumento prueba que $\alpha\beta$ es algebraico, considerando ahora

$$r(x) = \prod_{ij}(x-\alpha_i\alpha_j), \quad \text{y} \quad R(x, x_1, \ldots, x_n) = \prod_{ij}(x-x_ix_j).$$

Con esto ya tenemos probado que $\mathbb{A}$ es un anillo unitario, ya que, por ejemplo, si α es algebraico, como -1 también lo es, lo mismo sucede con $-\alpha$. Para concluir que forman un cuerpo sólo falta probar que si α es un número algebraico no nulo, entonces α^{-1} también es algebraico, pero eso es mucho más simple:

Si $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ es el polinomio mínimo de α, entonces $a_0 \neq 0$, pues de lo contrario tendría a x como factor y, al ser irreducible, tendría que ser meramente $p(x) = x$, pero entonces sería $\alpha = 0$. Tenemos que

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0,$$

luego

$$\alpha(\alpha^{n-1} + a_{n-1}\alpha^{n-2} + \cdots + a_1) = -a_0,$$

luego

$$\alpha^{-1} = \frac{1}{a_0}(\alpha^{n-1} + a_{n-1}\alpha^{n-2} + \cdots + a_1)$$

y concluimos que α^{-1} es algebraico porque se obtiene a partir de elementos de $\mathbb{A}$ (el propio α y números racionales) mediante sumas y productos. ∎

Como consecuencia, si K es un cuerpo obtenido a partir de $\mathbb{Q}$ mediante el teorema 6.1, al identificar sus elementos con números complejos, todos ellos son algebraicos, pues ciertamente α lo es (es raíz del polinomio dado $p(x)$) y todos los demás elementos de K se obtienen a partir de α y de números racionales mediante sumas y productos.

En particular todos los elementos de $\mathbb{Q}(i)$, o $\mathbb{Q}(\zeta)$, o $\mathbb{Q}\big(\sqrt{-2}\,\big)$, etc., son números algebraicos.

El cuerpo $\mathbb{A}$ cumple su propia versión del teorema fundamental del álgebra. Esto es consecuencia del teorema siguiente:

Teorema 8.10 *Si $p(x)$ es un polinomio con coeficientes en $\mathbb{A}$, sus raíces en $\mathbb{C}$ son números algebraicos.*

Demostración: Dividiendo entre el coeficiente director, podemos suponer que $p(x)$ es mónico. Pongamos que

$$p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_1 x + \alpha_0,$$

donde todos los coeficientes α_i son algebraicos. Sea $q(x)$ el producto de sus polinomios mínimos, que es un polinomio mónico con coeficientes racionales que tiene entre sus raíces a todos los α_i. Pongamos que

$$q(x) = (x - \beta_1)\cdots(x - \beta_m),$$

donde los α_i son algunos de los β_j. Consideremos el polinomio

$$P(x, x_0, \ldots, x_{n-1}) = x^n + x_{n-1}x^{n-1} + \cdots + x_1 x + x_0 \in \mathbb{Z}[x, x_0, \ldots, x_{n-1}]$$

y a su vez

$$R(x, y_1, \ldots, y_m) = \prod_{i_0,\ldots,i_{n-1}} P(x, y_{i_0}, \ldots, y_{i_{n-1}}) \in \mathbb{Z}[x, y_1, \ldots, y_m],$$

donde $i_0, \ldots, i_{n-1}$ toman todos los valores posibles entre 1 y m.

Visto como polinomio en $\mathbb{Z}[x][y_1,\dots,y_m]$, es simétrico, luego el teorema 8.5 nos da que

$$R(x,y_1,\dots,y_m) = g(x,e_1,\dots,e_m),$$

donde los polinomios $e_1,\dots,e_m$ son los polinomios simétricos elementales en $y_1,\dots,y_m$ y $g(x,t_1,\dots,t_m)$ es un polinomio en $\mathbb{Z}[x,t_1,\dots,t_m]$. Pero los números $a_i = e_i(\beta_1,\dots,\beta_m)$ son, salvo el signo, los coeficientes de $q(x)$, luego son números racionales, luego

$$h(x) = R(x,\beta_1,\dots,\beta_m) = g(x,a_1,\dots,a_m)$$

es un polinomio con coeficientes racionales y, por la definición de R como producto, uno de sus factores es precisamente $P(x,\alpha_0,\dots,\alpha_{n-1}) = P(x)$, luego una de las raíces de $h(x)$ es α. Esto prueba que α es algebraico. ∎

Por consiguiente:

Teorema 8.11 *Todo polinomio no constante con coeficientes en $\mathbb{A}$ tiene al menos una raíz en $\mathbb{A}$.*

Demostración: Por el teorema fundamental del álgebra, el polinomio tiene al menos una raíz en $\mathbb{C}$, y por el teorema anterior dicha raíz está en $\mathbb{A}$. ∎

Ejemplo Ahora es inmediato que el número real $\alpha = \sqrt[3]{5+\sqrt{2}}$ es algebraico, pues $\sqrt{2}$ lo es, ya que es raíz del polinomio x^2-2, y $5+\sqrt{2}$ también es algebraico, por ser suma de números algebraicos, y α también lo es por ser raíz del polinomio $x^3-(5+\sqrt{3})$, que tiene coeficientes algebraicos. ∎

Así, aunque identifiquemos con números complejos los elementos de los cuerpos como $\mathbb{Q}(i)$, etc., en realidad estamos trabajando, más concretamente, con elementos del cuerpo $\mathbb{A}$ de los números algebraicos, que es un cuerpo mucho menor.

Observemos además que si α es un número algebraico, lo mismo vale para su conjugado $\bar{\alpha}$, pues si se cumple que

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0,$$

donde los coeficientes son racionales, conjugando queda que

$$\bar{\alpha}^n + a_{n-1}\bar{\alpha}^{n-1} + \cdots + a_1\bar{\alpha} + a_0 = 0,$$

luego $\bar{\alpha}$ es raíz del mismo polinomio.

Como consecuencia, los elementos de $\mathbb{A}$ son los números complejos de la forma $z = x + yi$, donde x, y son números reales algebraicos.

En efecto, si x e y son algebraicos, como i también lo es, lo mismo vale para $x+yi$, mientras que si $z = x+yi$ es algebraico, lo mismo vale para su conjugado $\bar{z} = x-yi$, luego también para $2x = z+\bar{z}$, luego también para x, luego también para $yi = z-x$, luego también para $y = yi/i$.

Para terminar observemos que el teorema 6.1 aporta información no trivial sobre los números algebraicos:

Teorema 8.12 *Sea α un número algebraico de grado $n \geq 2$. Entonces*

1. *El conjunto $\mathbb{Q}(\alpha)$ formado por los números de la forma*

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1},$$

con $a_0, \ldots, a_{n-1} \in \mathbb{Q}$, es un cuerpo, y cada elemento de $\mathbb{Q}(\alpha)$ admite una única expresión de esta forma.

2. *Si K es cualquier cuerpo que contenga a $\mathbb{Q}$ en el que el polinomio $p(x)$ tenga una raíz α', existe una única aplicación $f : \mathbb{Q}(\alpha) \longrightarrow K$ que cumple:*

   (a) f permite identificar a $\mathbb{Q}(\alpha)$ con un subcuerpo de K, es decir, hace corresponder elementos distintos de $\mathbb{Q}(\alpha)$ con elementos distintos de K y además:

$$f(u+v) = f(u) + f(v), \qquad f(uv) = f(u)f(v).$$

   (b) $f(u) = u$ para todo número racional u.

   (c) $f(\alpha) = \alpha'$.

DEMOSTRACIÓN: Sea K' el anillo de clases de restos de $\mathbb{Q}(x)$ módulo el polinomio $p(x)$. El teorema 6.1 nos da que K' es un cuerpo cuyos elementos son de la forma descrita en 1) cambiando α por $\alpha' = \bar{x}$. El apartado 5) de dicho teorema nos da una aplicación $f : K' \longrightarrow \mathbb{A}$ que nos permite identificar K' con un subcuerpo de $\mathbb{A}$. Concretamente, al aplicar f a un elemento

$$a_0 + a_1\alpha' + \cdots + a_{n-1}\alpha'^{n-1}$$

obtenemos

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1},$$

luego dicho subcuerpo es precisamente el conjunto $\mathbb{Q}(\alpha)$ descrito en el enunciado. Esto hace que la propiedad 5) que por 6.1 se cumple para K', se cumple también para $\mathbb{Q}(\alpha)$, y esto es el apartado 2) del enunciado. ∎

Por ejemplo, el polinomio $p(x) = x^3 - 2$ es irreducible en $\mathbb{Q}[x]$, porque tiene grado 3 y no tiene raíces. En cambio, una raíz en $\mathbb{C}$ (o en $\mathbb{A}$) es $\sqrt[3]{2}$. El teorema anterior nos asegura que el conjunto

$$\mathbb{Q}(\sqrt[3]{2}) = \{a + b\sqrt[3]{2} + c\sqrt[3]{4} \mid a, b, c \in \mathbb{Q}\}$$

es un subcuerpo de $\mathbb{A}$, cosa que puede comprobarse explícitamente mediante muchos cálculos rutinarios, pero el teorema anterior los hace innecesarios.

Los cuerpos dados por el teorema anterior se llaman *cuerpos numéricos* y son el principal objeto de estudio de la teoría algebraica de números y su aritmética tiene implicaciones nada triviales sobre la aritmética de los números enteros, como ya hemos podido comprobar en varias ocasiones. A falta de una base algebraica sólida, en este libro podremos tratar poco más que los cuerpos cuadráticos, es decir, los obtenidos a partir de números algebraicos de grado 2. Dedicaremos el capítulo siguiente a estudiarlos con detalle.

8.4 Enteros algebraicos

Consideremos el cuerpo $\mathbb{Q}\big(\sqrt{5}\,\big)$, formado por los números $a+b\sqrt{5}$, donde a,b son números racionales. Podemos pensar en estos números como "números abstractos" dados por el teorema 6.1, o bien como números complejos algebraicos, teniendo en cuenta el teorema 8.12. Para las consideraciones que haremos en este libro, será irrelevante cómo queramos concebirlos.

Pero en los capítulos precedentes hemos visto que, en realidad, no nos interesan tanto estos cuerpos como sus correspondientes anillos de enteros, como $\mathbb{Z}[i]$ (el anillo de los enteros de Gauss) $\mathbb{Z}[\zeta]$ (el anillo de los enteros de Eisenstein) o $\mathbb{Z}\big[\sqrt{-2}\,\big]$, con cuya aritmética hemos obtenido diversos resultados sobre la aritmética de los números enteros.

Por ello resulta natural definir el anillo de enteros de $\mathbb{Q}\big(\sqrt{5}\,\big)$ como anillo $\mathbb{Z}\big[\sqrt{5}\,\big]$ formado por los números de la forma $a+b\sqrt{5}$, donde a,b son números enteros.

Ciertamente, podemos considerar este anillo, pero sería un error considerarlo por definición como el anillo de los enteros de $\mathbb{Q}\big(\sqrt{5}\,\big)$. Observemos que no tenemos ninguna definición general de "entero", sino que hemos dado definiciones *ad hoc* en cada caso particular.

Por ejemplo, si $\mathbb{Q}(i)$ está formado por los números $a+bi$ con $a,b\in\mathbb{Q}$, resulta "natural" llamar enteros de Gauss a los números de esta forma con $a,b\in\mathbb{Z}$, o si $\mathbb{Q}(\zeta)$ está formado por los números de la forma $a+b\zeta$ con $a,b\in\mathbb{Q}$, resulta "natural" llamar enteros de Eisenstein a los números de esta forma con $a,b\in\mathbb{Z}$, etc.

Sin embargo, esta "naturalidad" no es tan "natural" como parece, pues según ese patrón resulta tan "natural" considerar que $\mathbb{Z}[\zeta]$ es el anillo de enteros del cuerpo $\mathbb{Q}(\zeta)$ como considerar que $\mathbb{Z}\big[\sqrt{-3}\,\big]$ es el anillo de enteros del cuerpo $\mathbb{Q}\big(\sqrt{-3}\,\big)$, pero esto resulta contradictorio, ya que $\mathbb{Z}[\zeta]$ y $\mathbb{Z}\big[\sqrt{-3}\,\big]$ son dos anillos distintos, mientras que $\mathbb{Q}(\zeta)$ y $\mathbb{Q}\big(\sqrt{-3}\,\big)$ son el mismo cuerpo, por lo que tenemos dos definiciones "naturales" distintas del anillo de enteros de un mismo cuerpo.

En efecto, en la sección 6.2 hemos visto que podemos identificar

$$\zeta=\frac{-1+\sqrt{-3}}{2},$$

y esto hace que todo número de la forma $a+b\zeta$, con $a,b\in\mathbb{Q}$ pueda expresarse en la forma $a'+b'\sqrt{-3}$ y viceversa, pero el propio ζ es un ejemplo de número de la forma $a+b\zeta$ con $a,b\in\mathbb{Z}$, que no es de la forma $a'+b'\sqrt{-3}$ con $a',b'\in\mathbb{Z}$ (se tiene que $a'=-1/2, b'=1/2$).

Desde un punto de vista algebraico, tenemos que $\mathbb{Z}[\zeta]$ es un dominio euclídeo, por lo que disfruta de una aritmética análoga a la de los números enteros, mientras que $\mathbb{Z}\big[\sqrt{-3}\,\big]$ no tiene factorización única, lo que apunta a que la "definición correcta" del anillo de enteros del cuerpo $\mathbb{Q}(\zeta)=\mathbb{Q}\big(\sqrt{-3}\,\big)$ no es $\mathbb{Z}\big[\sqrt{-3}\,\big]$, sino $\mathbb{Z}[\zeta]$.

Vamos a ver que en el caso del cuerpo $\mathbb{Q}\big(\sqrt{5}\,\big)$ sucede algo similar, que la "definición correcta" de su anillo de enteros no es $\mathbb{Z}\big[\sqrt{5}\,\big]$, sino $\mathbb{Z}[\epsilon]$, donde

$$\epsilon = \frac{1+\sqrt{5}}{2}.$$

Para ello vamos a ver que existe una definición "natural" algebraica del concepto de anillo de enteros de cualquier cuerpo numérico:

Definición 8.13 Un número algebraico es un *entero algebraico* si es raíz de un polinomio mónico no nulo con coeficientes en $\mathbb{Z}$.

Por ejemplo, el número ϵ que acabamos de considerar es un entero algebraico de $\mathbb{Q}\big(\sqrt{5}\,\big)$, pues es raíz del polinomio $x^2 - x - 1$, e igualmente ζ es un entero algebraico, pues es raíz del polinomio $x^2 + x + 1$.

En cambio, el polinomio mínimo de

$$\alpha = \frac{1+\sqrt{-2}}{2}$$

es $x^2 - x + 3/4$, que no tiene coeficientes enteros. Ahora bien, en principio, esto no nos permite concluir que α no sea un entero algebraico, pues podría ser raíz de otro polinomio mónico con coeficientes enteros. Sin embargo, el teorema siguiente nos asegura que no es así:

Teorema 8.14 *Un número algebraico es un entero algebraico si y sólo si su polinomio mínimo tiene coeficientes enteros.*

Demostración: Sea α un número algebraico. Obviamente, si su polinomio mínimo tiene coeficientes enteros, entonces α es raíz de un polinomio mónico con coeficientes enteros, luego es un entero algebraico. Supongamos ahora que α es un entero algebraico y sea $f(x)$ un polinomio mónico no nulo con coeficientes enteros que tenga a α por raíz. Podemos tomarlo del menor grado posible. Basta probar que $f(x)$ es irreducible en $\mathbb{Q}[x]$, pues entonces, por el teorema 8.7, podremos concluir que $f(x)$ es el polinomio mínimo de α.

Supongamos que $f(x)$ no es irreducible en $\mathbb{Q}[x]$, de modo que $f(x) = u(x)v(x)$, donde ambos factores son polinomios no constantes de $\mathbb{Q}[x]$ (luego ambos tienen grado estrictamente menor que el de $f(x)$). Si llamamos c_1 y c_2 al producto de los denominadores de los coeficientes de $u(x)$ y $v(x)$, respectivamente, y $c = c_1c_2$, entonces, cambiando $u(x)$ y $v(x)$ por $c_1u(x)$ y $c_2v(x)$, respectivamente, tenemos que $cf(x) = u(x)v(x)$, donde ahora los factores tienen coeficientes enteros (aunque no son necesariamente mónicos) y el grado de ambos es estrictamente menor que el de $f(x)$. No perdemos generalidad si suponemos que $c > 0$.

Si $c \neq 1$, tomamos un primo $p \mid c$. Si llamamos $\bar{u}(x)$ y $\bar{v}(x)$ a los polinomios en $\mathbb{Z}_p[x]$ que resultan de sustituir los coeficientes de $u(x)$ y $v(x)$ por sus restos módulo p, tenemos que $\bar{u}(x)\bar{v}(x) = 0$, pero $\mathbb{Z}_p[x]$ es un dominio íntegro, luego uno de los dos factores tiene que ser nulo, y esto significa que p divide a todos los coeficientes de $u(x)$ o a todos los de $v(x)$.

Por lo tanto, en la igualdad $cf(x) = u(x)v(x)$ podemos simplificar p y pasar a una igualdad análoga en la que no cambian los grados de $u(x)$ y $v(x)$, pero en la que c tiene un divisor primo menos. Tras repetir este proceso un número finito de veces, llegamos a una expresión similar con $c = 1$, es decir, llegamos a que $f(x) = u(x)v(x)$, donde u y v son polinomios con coeficientes enteros cuyos grados son estrictamente menores que el de $f(x)$.

Pero el coeficiente director de $f(x)$ (que es 1) es el producto de los coeficientes directores de u y v, que son enteros, luego ambos tienen coeficiente director ± 1 y, cambiando el signo a ambos factores si es preciso, podemos suponer que $u(x)$ y $v(x)$ son mónicos, pero uno de los dos tiene que tener a α por raíz, y esto contradice la minimalidad del grado de $f(x)$. ∎

Por lo tanto, ya podemos asegurar que $(1 + \sqrt{-2})/2$ no es un entero algebraico, al contrario que $(1 + \sqrt{5})/2$.

Casi todos los teoremas sobre números algebraicos que hemos probado en la sección anterior admiten una versión para enteros algebraicos:

Teorema 8.15 *El conjunto $\mathbb{E}$ de todos los enteros algebraicos es un subanillo de $\mathbb{A}$.*

Demostración: La prueba del teorema 8.9 (excepto la parte final, en la que se demuestra que el inverso de un número algebraico es algebraico) es válida sin cambio alguno, con la única diferencia de que ahora el polinomio $p(x)$ tiene coeficientes enteros, por lo que los números $a_i = e_i(\alpha_1, \ldots, \alpha_n)$ no son meros números racionales, sino que son enteros y el polinomio $r(x)$ que obtenemos ahora es mónico y tiene coeficientes enteros.

La parte final falla porque no podemos asegurar que $1/a_0$ sea un entero algebraico (pero la prueba es válida si suponemos que el polinomio mínimo de α tiene término independiente $a_0 = \pm 1$). ∎

Teorema 8.16 *Si $p(x)$ es un polinomio mónico con coeficientes en $\mathbb{E}$, sus raíces en $\mathbb{C}$ son enteros algebraicos.*

Demostración: La prueba del teorema 8.10 vale sin cambio alguno, sin más que observar que ahora el polinomio $q(x)$ tiene coeficientes enteros, por lo que los a_i no son meros números racionales, sino enteros, lo que hace que el polinomio $h(x)$ sea mónico y con coeficientes enteros, luego α es entero algebraico. ∎

Ahora ya tenemos una definición general de "anillo de enteros":

Definición 8.17 Si K es un subcuerpo de $\mathbb{C}$, llamaremos *anillo de enteros de K* al conjunto $\mathcal{O}$ de los enteros algebraicos contenidos en K

Notemos que ciertamente es un anillo, pues al sumar y multiplicar elementos de $\mathcal{O}$ el resultado está en $\mathcal{O}$ porque está en K (al ser K un cuerpo) y es entero por el teorema 8.15.

El caso más simple es:

Teorema 8.18 *El anillo de enteros de $\mathbb{Q}$ es $\mathbb{Z}$.*

DEMOSTRACIÓN: Si r es un número racional, su polinomio mínimo es $x - r$ y tiene coeficientes enteros si y sólo si $r \in \mathbb{Z}$, luego los enteros de $\mathbb{Q}$ son los enteros ordinarios. ∎

Nota El teorema anterior afirma que los únicos enteros algebraicos que son números racionales son los enteros ordinarios. Precisamente por este hecho, en la teoría algebraica de números es costumbre llamar "enteros racionales" a los enteros ordinarios (porque son los enteros del cuerpo de los números racionales), y también es frecuente usar "entero" en el sentido más general de "entero algebraico", especificando "entero racional" cuando se quiere hacer referencia a los elementos de $\mathbb{Z}$. ∎

Ejemplo *El anillo $\mathbb{Z}[i]$ de los enteros de Gauss es precisamente el anillo de los enteros algebraicos del cuerpo $\mathbb{Q}(i)$.*

En efecto, como i es un entero algebraico (y los enteros usuales también lo son), el teorema 8.15 nos asegura que todos los enteros de Gauss son enteros algebraicos. Para probar que son todos podemos usar el teorema 2.28: Si $\alpha \in \mathbb{Q}(i)$ es un entero algebraico, como es raíz de un polinomio con coeficientes enteros, en particular en $\mathbb{Z}[i]$ y este anillo tiene factorización única (y su cuerpo de cocientes es $\mathbb{Q}(i)$) podemos concluir que $\alpha \in \mathbb{Z}[i]$, luego en $\mathbb{Q}(i)$ no hay más enteros algebraicos que los enteros de Gauss. ∎

Con el mismo argumento podemos probar que el anillo de enteros de $\mathbb{Q}(\zeta)$ es el anillo de los enteros de Eisenstein, o que el anillo de los enteros de $\mathbb{Q}\big(\sqrt{-2}\,\big)$ es $\mathbb{Z}\big[\sqrt{-2}\,\big]$. En el capítulo siguiente estudiaremos sistemáticamente los cuerpos cuadráticos y sus anillos de enteros.

Teorema 8.19 *Si K es un subcuerpo de $\mathbb{A}$ y $\mathcal{O}$ es su anillo de enteros, todo elemento de K es de la forma $\alpha = \beta/n$, donde $\beta \in \mathcal{O}$ y $n \in \mathbb{Z}$. En particular K es el cuerpo de cocientes de $\mathcal{O}$.*

DEMOSTRACIÓN: Si $\alpha \in K$, entonces α es raíz de un polinomio mónico con coeficientes racionales, pero también de un polinomio con coeficientes enteros, no necesariamente mónico. Digamos que

$$a_m\alpha^m + a_{m-1}\alpha^{m-1} + a_{m-2}\alpha^{m-2}\cdots + a_1\alpha + a_0 = 0.$$

Entonces, llamando $n = a_m \neq 0$ y multiplicando por n^{m-1}, obtenemos que

$$(n\alpha)^m + a_{m-1}(n\alpha)^{m-1} + a_{m-2}n(n\alpha)^{m-2} + \cdots + a_1n^{m-2}(n\alpha) + a_0m^{n-1} = 0,$$

por lo que $\beta = n\alpha$ es raíz de un polinomio mónico con coeficientes enteros, es decir, $\beta \in \mathcal{O}$. ∎

En particular, $\mathbb{A}$ es el cuerpo de cocientes de $\mathbb{E}$.

Por último observamos que el teorema 8.16 implica que los enteros algebraicos satisfacen la conclusión del teorema 2.28:

Teorema 8.20 *Si K es un subcuerpo de $\mathbb{A}$ y $\mathcal{O}$ es su anillo de enteros, todo elemento $\alpha \in K$ que es raíz de un polinomio mónico con coeficientes en $\mathcal{O}$ está en $\mathcal{O}$.*

Capítulo IX

Enteros cuadráticos

Tal y como señalábamos en el capítulo anterior, el estudio sistemático de los cuerpos numéricos requiere una base algebraica muy superior a la que tenemos a nuestro alcance, pero con las técnicas elementales que estamos manejando podemos, no obstante, estudiar sistemáticamente los cuerpos cuadráticos y sus anillos de enteros, y de este estudio obtendremos, tanto en este capítulo como en los capítulos siguientes, numerosas consecuencias sobre la aritmética de los enteros racionales.

9.1 Cuerpos cuadráticos

Definición 9.1 Sea $d \neq 1$ un número entero libre de cuadrados. Definimos $\sqrt{d}$ como la única raíz cuadrada positiva de d si $d > 0$, o bien $\sqrt{d} = \sqrt{|d|}i$ si $d < 0$, es decir, como la única raíz cuadrada compleja de d con parte imaginaria positiva.

En cualquier caso $\sqrt{d}$ es un número algebraico de grado 2, cuyo polinomio mínimo es $x^2 - d$. El teorema 8.12 nos asegura que el conjunto $\mathbb{Q}\big(\sqrt{d}\,\big)$ de los números complejos de la forma $a + b\sqrt{d}$, con $a, b \in \mathbb{Q}$, es un cuerpo de números algebraicos.

Más aún, puesto que $-\sqrt{d}$ también es raíz del polinomio $x^2 - d$, dicho teorema nos asegura que la conjugación $\mathbb{Q}\big(\sqrt{d}\,\big) \longrightarrow \mathbb{Q}\big(\sqrt{d}\,\big)$ dada por

$$\overline{a + b\sqrt{d}} = a - b\sqrt{d}$$

cumple las relaciones

$$\overline{\alpha + \beta} = \bar{\alpha} + \bar{\beta}, \qquad \overline{\alpha\beta} = \bar{\alpha}\,\bar{\beta}.$$

A su vez, la conjugación nos permite definir la norma $\mathrm{N} : \mathbb{Q}\big(\sqrt{d}\,\big) \longrightarrow \mathbb{Q}$ mediante $\mathrm{N}(\alpha) = \alpha\bar{\alpha}$ o, más explícitamente,

$$\mathrm{N}(a + b\sqrt{d}) = (a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - db^2,$$

y tiene la propiedad de que $\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta)$.

Notemos que $\mathrm{N}(\alpha) = 0$ sólo sucede cuando $\alpha = 0$, pues equivale a que $\alpha\bar{\alpha} = 0$ y, como $\mathbb{Q}(\sqrt{d})$ es un cuerpo, esto equivale a que $\alpha = 0$ o $\bar{\alpha} = 0$, pero en ambos casos $\alpha = 0$.

Por lo tanto, el inverso de un elemento no nulo de $\mathbb{Q}(\sqrt{d})$ puede calcularse mediante la fórmula habitual:

$$\alpha^{-1} = \frac{\bar{\alpha}}{\mathrm{N}(\alpha)}.$$

Nota Observemos que para que el polinomio $x^2 - n$ sea irreducible no hace falta que n sea libre de cuadrados, sino que basta con que n no sea un cuadrado perfecto, luego el cuerpo $\mathbb{Q}(\sqrt{n})$ está definido bajo esta hipótesis más general, pero en realidad esta generalidad es aparente, pues podemos expresar $n = m^2 d$, donde d es libre de cuadrados, y entonces $\sqrt{n} = m\sqrt{d}$, de donde se sigue inmediatamente que los números de la forma $a + b\sqrt{n}$, con $a, b \in \mathbb{Q}$ son los mismos que los de la forma $a' + b'\sqrt{d}$, con $a', b' \in \mathbb{Q}$, es decir, que $\mathbb{Q}(\sqrt{n})$ y $\mathbb{Q}(\sqrt{d})$ son en realidad el mismo cuerpo cuadrático.

Más en general, si $ax^2 + bx + c$ es cualquier polinomio irreducible en $\mathbb{Q}[x]$, un razonamiento similar muestra que el cuerpo dado por 8.12 coincide con $\mathbb{Q}(\sqrt{D})$, donde $D = b^2 - 4ac$, que a su vez coincide con el cuerpo $\mathbb{Q}(\sqrt{d})$ determinado por la relación $D = m^2 d$, con d libre de cuadrados. Así pues, los cuerpos $\mathbb{Q}(\sqrt{d})$ que hemos definido son una familia de cuerpos bastante más general de lo que podría parecer en un principio, pues son todos los cuerpos cuadráticos, es decir, todos los cuerpos obtenidos por el teorema 8.12 a partir de un polinomio de segundo grado irreducible en $\mathbb{Q}[x]$. ∎

Cuerpos cuadráticos reales e imaginarios Según acabamos de indicar, los cuerpos de la forma $\mathbb{Q}(\sqrt{d})$ se conocen como *cuerpos cuadráticos*. Los correspondientes a $d > 0$ se llaman *cuerpos cuadráticos reales,* mientras que los correspondientes a $d < 0$ se llaman *imaginarios.*

La razón de esta nomenclatura es evidente: los que llamamos cuerpos cuadráticos reales son, más concretamente, subcuerpos de $\mathbb{R}$, es decir, que están formados exclusivamente por números reales, mientras que los que llamamos cuerpos cuadráticos imaginarios contienen tanto números reales como números imaginarios.

Observemos que la inclusión $\mathbb{Q}(\sqrt{d}) \subset \mathbb{R}$ nos aporta una información relevante sobre los cuerpos cuadráticos reales, y es que en ellos es posible definir una relación de orden que los convierte en cuerpos ordenados y respecto a la cual $\sqrt{d} > 0$.

Esta relación puede calcularse en la práctica mediante manipulaciones puramente algebraicas. Por ejemplo, si queremos saber si

$$3 - 2\sqrt{2} < \frac{\sqrt{2}}{4}$$

sólo tenemos que operar con las reglas válidas en todo cuerpo ordenado, según las cuales esto equivale a $12 - 8\sqrt{2} < \sqrt{2}$, o a $12 < 9\sqrt{2}$, o a $12^2 < 9^2 \cdot 2$, es decir, a $144 < 162$, lo cual es cierto.

En cambio, los cuerpos cuadráticos imaginarios no admiten una relación de orden que los convierta en cuerpos ordenados, porque en un cuerpo ordenado los números negativos no pueden tener raíz cuadrada.

Otra diferencia es que la norma $\mathrm{N}(a + b\sqrt{d}) = a^2 + db^2$ no toma valores negativos en los cuerpos cuadráticos reales, pero sí en los imaginarios.

Una última observación es que, en un cuerpo cuadrático imaginario $\mathbb{Q}\big(\sqrt{d}\,\big)$, la conjugación $a+b\sqrt{d} \mapsto a-b\sqrt{d}$ que hemos definido no es más que la restricción de la conjugación compleja (pues el conjugado complejo de $\sqrt{d} = \sqrt{|d|}i$ es $-\sqrt{d}$), mientras que en los cuerpos cuadráticos reales la conjugación no tiene nada que ver con la conjugación compleja, que deja invariantes a todos los números reales. ∎

Si $K = \mathbb{Q}\big(\sqrt{d}\,\big)$ es un cuerpo cuadrático y $\alpha = a + b\sqrt{d}$ es uno de sus elementos, sabemos que α es un número algebraico, y es fácil determinar su polinomio mínimo. Si $b = 0$, obviamente $\operatorname{pol\,mín}\alpha = x - a$, mientras que si $b \neq 0$, entonces

$$\operatorname{pol\,mín}\alpha = (x - \alpha)(x - \bar{\alpha}) = x^2 - 2a + \mathrm{N}(\alpha),$$

pues este polinomio es mónico, tiene coeficientes racionales, tiene a α por raíz y, como α es irracional, no tiene raíces en $\mathbb{Q}$, luego es irreducible en $\mathbb{Q}[x]$.

Ahora es fácil determinar los enteros algebraicos de $\mathbb{Q}\big(\sqrt{d}\,\big)$. Recordemos que, por el teorema 8.14, un número algebraico es entero si y sólo si su polinomio mínimo tiene coeficientes enteros. Por lo tanto, un número $\alpha = a+b\sqrt{d}$ será un entero algebraico si $b = 0$ y $a \in \mathbb{Z}$, o, en caso de que $b \neq 0$, lo será si y sólo si

$$2a \in \mathbb{Z}, \qquad c = \mathrm{N}(\alpha) = a^2 - db^2 \in \mathbb{Z}.$$

Tenemos que $4c = (2a)^2 - (2b)^2 d$, luego

$$(2b)^2 d = 4c - (2a)^2$$

es entero. De aquí se sigue que $2b$ es entero. En efecto, si $2b = u/v$, donde u y v son enteros primos entre sí, tenemos que $u^2 d/v^2$ es entero, luego $v^2 \mid d$ y, como d es libre de cuadrados, $v = \pm 1$.

Así pues, $x = 2a$ e $y = 2b$ son enteros (pero no necesariamente pares, pues a y b no son necesariamente enteros). Si $d \equiv 2, 3 \pmod 4$, la congruencia

$$x^2 - y^2 d \equiv 0 \pmod 4$$

sólo puede satisfacerse si x e y son pares (pues los cuadrados módulo 4 son 0 y 1), luego a y b son enteros.

Si $d \equiv 1$ (mód 4) (notemos que $d \equiv 0$ (mód 4) es imposible, ya que d es libre de cuadrados), además de en el caso en que x e y sean pares, la congruencia se satisface también si x e y son ambos impares, es decir, si $x \equiv y$ (mód 2). Entonces

$$\alpha = \frac{x}{2} + \frac{y}{2}\sqrt{d} = \frac{x-y}{2} + y\frac{1+\sqrt{d}}{2}.$$

Con esto casi tenemos demostrado el teorema siguiente:

Teorema 9.2 *Sea $d \neq 1$ un entero libre de cuadrados. Entonces el anillo de enteros de $\mathbb{Q}(\sqrt{d})$ es $\mathbb{Z}[\omega]$, decir, el conjunto formado por los elementos de la forma $a + b\omega$, donde $a, b \in \mathbb{Z}$ y*

$$\omega = \begin{cases} \sqrt{d} & \text{si } d \not\equiv 1 \text{ (mód 4)}, \\ \dfrac{1+\sqrt{d}}{2} & \text{si } d \equiv 1 \text{ (mód 4)}. \end{cases}$$

Demostración: Hemos probado que todo entero algebraico tiene que ser de la forma indicada en el enunciado, pero falta probar que los elementos de esta forma son realmente enteros algebraicos, para lo cual basta probar que ω lo es. Esto es inmediato cuando $\omega = \sqrt{d}$ y, en el caso alternativo, basta observar que

$$\operatorname{pol\,mín}\omega = x^2 - x + \frac{1-d}{4}.$$

∎

Notemos que en el caso $d = -3$ hemos definido

$$\omega = \frac{1+\sqrt{-3}}{2}, \qquad \zeta = \frac{-1+\sqrt{-3}}{2}.$$

La relación entre ambos es $\omega = \zeta + 1 = -\zeta^2$ y es obvio que $\mathbb{Z}[\omega] = \mathbb{Z}[\zeta]$, pues un elemento de $\mathbb{Q}(\sqrt{-3})$ puede expresarse en la forma $a + b\omega$, con a y b enteros, si y sólo si puede expresarse en la forma $a + b\zeta$, donde a y b son (otros) enteros.

Si $K = \mathbb{Q}(\sqrt{d})$ es un cuerpo cuadrático y $\mathbb{Z}[\omega]$ es su anillo de enteros, la norma se resgringe a una aplicación $\mathrm{N} : \mathbb{Z}[\omega] \longrightarrow \mathbb{Z}$.

En efecto, la norma de un entero algebraico es un entero racional, porque es ciertamente es racional, y además es entera (es el producto de un entero y su conjugado, que también es entero).

9.2 Grupos de unidades

El primer paso a la hora de estudiar la aritmética de un anillo es conocer sus unidades. En general, si k es un cuerpo cuadrático, es habitual referir a k los conceptos que, en sentido estricto, dependen de su anillo de enteros. Por ejemplo, es habitual hablar de "las unidades de k" para referirse a las unidades del anillo de enteros de k. Así hay que entender, por ejemplo, el enunciado del teorema siguiente:

Teorema 9.3 *Las unidades de un cuerpo cuadrático son los enteros ϵ que cumplen* $\mathrm{N}(\epsilon) = \pm 1$.

DEMOSTRACIÓN: Si ϵ es entero y $\mathrm{N}(\epsilon) = \pm 1$, entonces $\epsilon(\pm\bar{\epsilon}) = 1$, luego existe $\epsilon^{-1} = \pm\bar{\epsilon}$ en $\mathbb{Z}[\omega]$ y ϵ es una unidad. Recíprocamente, si ϵ es una unidad, tenemos que $\epsilon\epsilon^{-1} = 1$, luego $\mathrm{N}(\epsilon)\,\mathrm{N}(\epsilon^{-1}) = \mathrm{N}(1) = 1$, y como los dos factores son enteros, tiene que ser $\mathrm{N}(\epsilon) = \pm 1$. ∎

Esto nos permite identificar las unidades de todos cuerpos cuadráticos imaginarios. El teorema siguiente muestra que los casos que ya hemos estudiado de los enteros de Gauss y los enteros de Eisenstein son casos excepcionales entre los cuerpos cuadráticos imaginarios en lo tocante a sus grupos de unidades:

Teorema 9.4 *Si $K = \mathbb{Q}\big(\sqrt{d}\,\big)$ es un cuerpo cuadrático imaginario, sus unidades son ± 1 excepto si $d = -1$ o $d = -3$.*

- *Si $d = -1$ las unidades son ± 1 y $\pm i$ (donde $i = \sqrt{-1} = \omega$).*
- *Si $d = -3$ las unidades son $\pm 1, \pm\omega, \pm\omega^2$.*

DEMOSTRACIÓN: Como $d < 0$, la norma es siempre positiva y hemos visto que un entero $a + b\sqrt{d}$ es una unidad si y sólo si $a^2 - db^2 = 1$. Los casos $d = -1$ y $d = -3$ ya los hemos estudiado, pero repetimos el argumento:

Para $d = -1$ la ecuación se reduce a $a^2 + b^2 = 1$, y además los enteros de $\mathbb{Q}\big(\sqrt{-1}\,\big)$ son los elementos de $\mathbb{Z}\big[\sqrt{-1}\,\big]$, luego a y b han de ser enteros racionales. Es claro que las únicas soluciones enteras de $a^2 + b^2 = 1$ son $(1, 0)$, $(0, 1)$, $(-1, 0)$ y $(0, -1)$, de donde las unidades son las indicadas.

Para $d = -2$ es claro que la ecuación $a^2 + 2b^2$ sólo tiene las soluciones $(\pm 1, 0)$, que se corresponden con las unidades ± 1.

Para $d = -3$ tenemos que el anillo de enteros es $\mathbb{Z}[\omega]$ y las unidades cumplen

$$\mathrm{N}(a + b\omega) = (a + b\omega)(a + b\bar{\omega}) = a^2 + ab + b^2 = 1.$$

Equivalentemente:

$$\frac{1}{4}((2a)^2 + 2(2a)b + b^2 + 3b^2) = \frac{1}{4}((2a + b)^2 + 3b^2) = 1.$$

A partir de esta expresión es fácil ver que los únicos valores posibles para (a, b) son $(\pm 1, 0), (0, \pm 1), (\pm 1, \mp 1)$, que da lugar a las unidades

$$\pm 1, \quad \pm\omega, \quad \pm(\omega - 1) = \pm\omega^2.$$

Si $d < -3$ hemos de tener presente que a y b pueden ser enteros o semienteros, es decir, $a = A/2$, $b = B/2$, con A, $B \in \mathbb{Z}$. Como el caso semientero incluye al caso entero, basta probar que las únicas soluciones semienteras de $a^2 - db^2 = 1$ son $(1, 0)$ y $(-1, 0)$. En efecto, tenemos $A^2 - dB^2 = 4$, pero si $B \neq 0$, entonces $A^2 - dB^2 > 4$, pues en realidad $d \leq -5$, luego ha de ser $B = 0$ y $A^2 = 4$, o sea, $a = \pm 1$, $b = 0$. ∎

Observemos que, en el caso $d = -3$, la relación $\omega = -\zeta^2$ implica que las seis unidades de $\mathbb{Z}[\omega] = \mathbb{Z}[\zeta]$ pueden expresarse también como $\pm 1, \pm\zeta, \pm\zeta^2$, tal y como habíamos visto al estudiar los enteros de Eisenstein.

La situación en el caso de los cuerpos cuadráticos reales es muy diferente. Examinemos la situación en algunos casos concretos. Según el teorema 9.3, las unidades en $\mathbb{Q}\big(\sqrt{2}\,\big)$ son los enteros $\epsilon = x + y\sqrt{2}$ tales que $\mathrm{N}(\epsilon) = x^2 - 2y^2 = \pm 1$. Si, con la ayuda de un ordenador, damos a x, y todos los valores posibles entre 0 y 100, obtenemos las unidades que muestran las tablas siguientes, en las que hemos incluido el resultado de búsquedas análogas para $\mathbb{Q}\big(\sqrt{3}\,\big)$, $\mathbb{Q}\big(\sqrt{5}\,\big)$ y $\mathbb{Q}\big(\sqrt{6}\,\big)$. Hemos incluido los datos de $\mathbb{Q}\big(\sqrt{5}\,\big)$ en una tabla separada porque aparecen muchas más unidades. Notemos que en este caso $\mathrm{N}(x + y\omega) = x^2 + xy - y^2$.

	0	1	2	3	4	5	6
ϵ	1	$1+\sqrt{2}$	$3+2\sqrt{2}$	$7+5\sqrt{2}$	$17+12\sqrt{2}$	$41+29\sqrt{2}$	$99+70\sqrt{2}$
x^2-2y^2	1	-1	1	-1	1	-1	1
ϵ	1	$2+\sqrt{3}$	$7+4\sqrt{3}$	$26+15\sqrt{3}$	$97+56\sqrt{3}$		
x^2-3y^2	1	1	1	1	1		
ϵ	1	$5+2\sqrt{6}$	$49+20\sqrt{6}$				
x^2-6y^2	1	1	1				

$\mathbb{Q}\big(\sqrt{5}\,\big)$	0	1	2	3	4	5	6
ϵ	1	ω	$1+\omega$	$1+2\omega$	$2+3\omega$	$3+5\omega$	$5+8\omega$
x^2+xy-y^2	1	-1	1	-1	1	-1	

	7	8	9	10	11
ϵ	$8+13\omega$	$13+21\omega$	$21+34\omega$	$34+55\omega$	$55+89\omega$
x^2+xy-y^2	-1	1	-1	1	-1

Si pensáramos en los valores de la tabla como meras soluciones enteras (x, y) de ecuaciones como $x^2 - 2y^2 = \pm 1$, no es fácil en muchos casos encontrar un patrón —que lo hay— aunque éste es especialmente simple en el caso $d = 5$. Ahora bien, si tenemos en cuenta que lo que estamos calculando son unidades de un anillo, hay una forma obvia de generar nuevas soluciones a partir de una dada, y es que si ϵ es una unidad, sabemos que ϵ^n es una unidad para todo exponente entero n.

Pero, si llamamos η a la segunda unidad de cada tabla (la situada en la columna etiquetada con un 1, es decir, descartando $\epsilon = 1$), vemos que las unidades de cada tabla no son sino $\eta^0 = 1, \eta^1, \eta^2, \ldots$

Ejercicio: Calcular más potencias de la unidad η hasta llenar los huecos que quedan en las tablas.

Si calculamos las potencias con exponente negativo no obtenemos nada esencialmente nuevo. Por ejemplo, para $d = 2$ obtenemos

$$\eta^{-1} = -1 + \sqrt{2}, \quad \eta^{-2} = 3 - 2\sqrt{2}, \quad \eta^{-3} = 7 - 5\sqrt{2}, \ldots$$

Esto es fácil de interpretar: como $\mathrm{N}(\eta) = \eta\bar{\eta} = -1$, resulta que $\eta^{-1} = -\bar{\eta}$ y, en general, $\eta^{-n} = (-1)^n\overline{\eta^n}$, luego η^n y η^{-n} se diferencian únicamente en un signo.

Es obvio que todas las unidades η^n de las tablas (excepto $\eta^0 = 1$) cumplen $\eta^n > 1$, luego las potencias con exponente negativo cumplen $0 < \eta^{-n} < 1$. A estas unidades hay que añadir las de la forma $-\eta^n$, que son las unidades negativas.

No tenemos ningún indicio de que haya más, así que resulta natural conjeturar que todas las unidades de un cuerpo cuadrático real pueden expresarse en la forma $\pm\eta^n$, donde el exponente varía en $\mathbb{Z}$, para una cierta unidad "fundamental" fija η.

La mayor parte de esta conjetura se puede justificar mediante una serie de razonamientos relativamente simples. Lo único que no es evidente en absoluto es la mera existencia de unidades no triviales:

En todo cuerpo cuadrático real existe una unidad distinta de ± 1.

Puesto que justificar este hecho nos va a llevar a consideraciones un poco alejadas de las técnicas que hemos empleado hasta ahora, vamos a admitir de momento que es así y vamos a extraer las consecuencias básicas que ello conlleva.

Consideramos, pues, un cuerpo cuadrático real arbitrario $k = \mathbb{Q}(\sqrt{d})$ y suponemos que su anillo de enteros $\mathbb{Z}[\omega]$ contiene al menos una unidad $\epsilon \neq \pm 1$.

En primer lugar observamos que, cambiando ϵ por $-\epsilon$ si es preciso, podemos tomar $\epsilon > 0$ y, cambiando ϵ por $1/\epsilon$, podemos tomar $\epsilon > 1$. Vamos a estudiar con más detalle las unidades $\epsilon > 1$. El teorema siguiente afirma que al buscar unidades $\epsilon = x + y\omega$ con $x, y \geq 0$, como hemos hecho para construir las tablas precedentes, nos encontraremos de hecho con todas las unidades $\epsilon \geq 1$:

Teorema 9.5 *Si $\epsilon > 1$ es una unidad del anillo de enteros $\mathbb{Z}[\omega]$ de un cuerpo cuadrático real $\mathbb{Q}(\sqrt{d})$, entonces $\epsilon = a + b\omega$ con $a, b \geq 1$, excepto si $d = 5$ y $\epsilon = \omega$ (en cuyo caso $a = 0$ y $b = 1$).*

Demostración: Ciertamente, podemos expresar $\epsilon = a + b\omega$, donde a y b son enteros racionales. Como $\mathrm{N}(\epsilon) = \epsilon\bar{\epsilon} = \pm 1$, tenemos que $\bar{\epsilon} = \pm 1/\epsilon$, donde $0 < 1/\epsilon < 1$, luego, cualquiera que sea el signo, $\epsilon - \bar{\epsilon} > 0$. Esto equivale a que $b(\omega - \bar{\omega}) > 0$, pero

$$\omega - \bar{\omega} = \begin{cases} 2\sqrt{d} & \text{si } d \not\equiv 1 \pmod 4, \\ \sqrt{d} & \text{si } d \equiv 1 \pmod 4, \end{cases}$$

luego en cualquier caso $\omega - \bar{\omega} > 0$, luego $b > 0$.

Por otra parte, $\bar{\omega} < -1$ salvo si $d = 5$. En efecto, si $d \not\equiv 1 \pmod 4$ es $\bar{\omega} = -\sqrt{d} < -1$, ya que $d \geq 2$, y si $d \equiv 1 \pmod 4$ entonces la desigualdad

$$\bar{\omega} = \frac{1 - \sqrt{d}}{2} < -1$$

equivale a que $\sqrt{d} > 3$ o a que $d > 9$, lo cual sucede salvo si $d = 5$. Por lo tanto, salvo en este caso excepcional, para que se cumpla

$$|a + b\bar{\omega}| = |\bar{\epsilon}| < 1$$

tiene que ser $a > 0$.

En el caso excepcional podemos probar de todos modos que $a \geq 0$. En efecto, tenemos que

$$-1 < \bar{\epsilon} = a + b\bar{\omega} < 1,$$

luego

$$a > -1 - b\bar{\omega} \geq -1 - \bar{\omega} = -1 - \frac{1-\sqrt{5}}{2} \approx -0.38,$$

luego $a \geq 0$. Pero si $a = 0$, entonces $\epsilon = b\omega$ y $\mathrm{N}(\epsilon) = -b^2$, luego necesariamente $b = \pm 1$, pero como $\epsilon > 0$, tiene que ser $b = 1$, y así la única excepción es $\epsilon = \omega$.

■

Ahora es fácil ver que todas las unidades son, salvo signo, potencias de una unidad fija:

Teorema 9.6 *Si $\mathbb{Q}(\sqrt{d})$ es un cuerpo cuadrático real, su anillo de enteros contiene una unidad $\eta > 1$ tal que cada unidad se expresa de forma única como $\epsilon = \pm\eta^n$, para un cierto entero racional n.*

Demostración: Sabemos que existe una unidad $\epsilon > 1$. En virtud del teorema anterior, si $1 < a' + b'\omega < a + b\omega$ son dos unidades, o bien $0 \leq a' < a$ o bien $0 < b' < b$, lo que significa que por debajo de una unidad ϵ hay a lo sumo un número finito de unidades $\epsilon' > 1$. En efecto, para cada $0 \leq a' < a$ podemos estudiar si la ecuación $\mathrm{N}(a' + b'\omega) = \pm 1$ tiene solución para algún $b' > 0$ (es una ecuación cuadrática, luego a lo sumo habrá dos valores de b' para cada signo del término independiente) e igualmente si para cada $0 < b' < b$ la ecuación tiene solución para algún $a' \geq 0$, luego el número de posibilidades es finito.

Por consiguiente, existe una unidad $\eta > 1$ mínima en el anillo de enteros de $\mathbb{Q}(\sqrt{d})$. Vamos a comprobar que cumple lo requerido. Teniendo en cuenta que (por el binomio de Newton)

$$\eta^m = (1 + \eta - 1)^m > 1 + m(\eta - 1)$$

si $\epsilon > 1$ es cualquier otra unidad, existirá un mínimo número natural m tal que $\epsilon < \eta^m$. Sea $n = m - 1$, de modo que $\eta^n \leq \epsilon < \eta^{n+1}$, luego $1 \leq \epsilon/\eta^n < \eta$, pero ϵ/η^n es una unidad, luego la minimalidad de η implica que $1 = \epsilon/\eta^n$, es decir, que $\epsilon = \eta^n$. Necesariamente $n \geq 1$ y la expresión es única, pues si $n < n'$ entonces $\eta^n < \eta^{n'}$.

Si $0 < \epsilon < 1$ aplicamos la parte ya probada a ϵ^{-1}, que será de la forma $\epsilon^{-1} = \eta^n$, luego $\epsilon = \eta^{-n}$.

Con esto hemos probado que toda unidad $\epsilon > 0$ es de la forma $\epsilon = \eta^n$ para un único entero n (positivo si $\epsilon > 1$, nulo si $\epsilon = 1$ y negativo si $\epsilon < 1$). Finalmente, si $\epsilon < 0$, entonces al aplicar la parte ya probada a $-\epsilon$ obtenemos que $\epsilon = -\eta^n$.

■

Definición 9.7 Una *unidad fundamental* de un cuerpo cuadrático real $\mathbb{Q}(\sqrt{d})$ es una unidad η de su anillo de enteros tal que cualquier otra unidad se expresa de forma única como $\epsilon = \pm\eta^n$, donde n es un entero racional.

Salvo que tenemos pendiente justificar la existencia de unidades no triviales, hemos probado que todo cuerpo cuadrático real tiene una unidad fundamental $\eta > 1$. Es fácil ver que es única, pues si $\eta' > 1$ fuera otra, tendría que ser $\eta = \eta'^n$ y $\eta' = \eta^m$, para ciertos exponentes naturales no nulos, pero entonces $\eta^{nm} = \eta$, lo cual sólo es posible si $nm = 1$, es decir, si $n = m = 1$, luego $\eta = \eta'$.

Similarmente se comprueba que hay exactamente cuatro unidades fundamentales, que son $\pm\eta$ y $\pm\eta^{-1}$. No obstante, cuando hablemos de "la unidad fundamental" de un cuerpo cuadrático, nos referiremos a la que cumple $\eta > 1$.

Las tablas que hemos calculado nos permiten identificar fácilmente la unidad fundamental de los cuerpos cuadráticos considerados en ellas:

Ejemplo *La unidad fundamental del anillo de enteros de* $\mathbb{Q}(\sqrt{5})$ *es* ω.

En efecto, si no fuera la unidad fundamental, habría otra unidad

$$1 < a + b\omega < \omega,$$

con $a, b \geq 1$, según el teorema 9.5, pero entonces $\omega \leq b\omega \leq a + b\omega < \omega$, con lo que tenemos una contradicción.

Si calculamos las potencias sucesivas de ω obtenemos:

n	1	2	3	4	5	6
ω^n	ω	$1+\omega$	$1+2\omega$	$2+3\omega$	$3+5\omega$	$5+8\omega$

En general, es fácil ver que si $\omega^n = x_n + y_n\omega$, se cumple que

$$(x_1, y_1) = (0, 1), \qquad (x_{n+1}, y_{n+1}) = (y_n, x_n + y_n).$$

Es fácil ver entonces que, salvo por $y_1 = y_2$ y $x_2 = x_3$, las sucesiones x_n e y_n son estrictamente crecientes. ∎

Para identificar unidades fundamentales resulta útil comprobar algo que se observa en las tablas que hemos calculado:

Teorema 9.8 *Sea* $\eta > 1$ *la unidad fundamental del cuerpo* $\mathbb{Q}(\sqrt{d})$ *y, para cada* $n \geq 1$*, sea* $\eta^n = x_n + y_n\omega$*. Entonces las sucesiones* x_n *e* y_n *son estrictamente crecientes salvo si* $d = 5$.

Demostración: Supongamos en primer lugar que $d \not\equiv 1 \pmod 4$. Entonces $\omega = \sqrt{d}$ y $\eta = a + b\sqrt{d}$ con $a, b > 0$. Basta tener en cuenta que

$$x_{n+1} + y_{n+1}\sqrt{d} = (x_n + y_n\sqrt{d})(a + b\sqrt{d}) = x_n a + y_n bd + (x_n b + y_n a)\sqrt{d},$$

de modo que, si $x_n, y_n > 0$, entonces

$$x_{n+1} = x_n a + y_n bd > x_n > 0, \qquad y_{n+1} = x_n b + y_n a > y_n > 0.$$

Consideremos ahora el caso en que $d \equiv 1 \pmod 4$. Entonces $\omega = (1+\sqrt{d})/2$ y cumple

$$\omega^2 = \omega + h,$$

donde $h = (d-1)/4 \geq 1$.

Por lo tanto, Si $\eta = a + b\omega$,

$$\begin{aligned} x_{n+1} + y_{n+1}\omega &= (x_n + y_n\omega)(a + b\omega) = x_n a + y_n b\omega^2 + (x_n b + y_n a)\omega \\ &= x_n a + y_n bh + (x_n b + y_n a + y_n b)\omega. \end{aligned}$$

Sabemos que si $d \neq 5$, se cumple que $a, b \geq 1$, luego podemos concluir igual que en el caso anterior. ∎

Esto significa que, para encontrar la unidad fundamental de un cuerpo cuadrático real $\mathbb{Q}\big(\sqrt{d}\big)$, basta encontrar una unidad cualquiera $\epsilon = a + b\omega$. Cambiándola por $-\epsilon$ o por $\pm\epsilon^{-1}$ podemos hacer que sea $\epsilon > 1$, y ahora basta estudiar si alguno de los números $x + y\omega$ con $0 < x < a$, $0 < y < b$ es una unidad. El menor de todos ellos será la unidad fundamental.

Ejemplo Consideremos el cuerpo $\mathbb{Q}\big(\sqrt{2}\big)$. Su anillo de enteros es $\mathbb{Z}\big[\sqrt{2}\big]$ y las unidades $\epsilon = u + v\sqrt{2}$ deben cumplir $u^2 - 2v^2 = \pm 1$. Obviamente una unidad no trivial es $\eta = 1 + \sqrt{2}$, y es la unidad fundamental, pues otra unidad $1 < \epsilon = u + v\sqrt{2} < \eta$ debe cumplir $0 < u < 1$ y $0 < v < 1$, lo cual es imposible. Otros ejemplos de unidades son

$$\eta^2 = 3 + 2\sqrt{2}, \quad \eta^3 = 7 + 5\sqrt{2}, \quad \eta^4 = 17 + 12\sqrt{2}, \ldots$$

las unidades $\pm\eta^{-n}$ sólo se diferencian de éstas en los signos de las coordenadas. Teniendo en cuenta que $\mathrm{N}(\eta) = -1$, a partir de aquí es fácil concluir que las soluciones enteras de la ecuación

$$x^2 - 2y^2 = -1$$

son, salvo signos, las asociadas a las potencias η^{2n+1} con $n \geq 0$, es decir, los pares (x, y) tales que

$$x + y\sqrt{2} = (1 + \sqrt{2})(3 + 2\sqrt{2})^n, \qquad n = 0, 1, 2, \ldots$$

Por otro lado, las soluciones de la ecuación $x^2 - 2y^2 = 1$ son las asociadas a potencias η^{2n}, es decir, los pares (x, y) tales que

$$x + y\sqrt{2} = (3 + 2\sqrt{2})^n, \qquad n = 0, 1, 2, \ldots$$

∎

Ejercicio: Encontrar expresiones para las soluciones enteras de las ecuaciones

$$x^2 - xy + y^2 = -1, \quad x^2 - xy + y^2 = 1, \quad x^2 - 5y^2 = -1, \quad x^2 - 5y^2 = 1.$$

Ejercicio: Encontrar la unidad fundamental del anillo de enteros del cuerpo $\mathbb{Q}(\sqrt{3})$. Observar que $\mathrm{N}(\eta) = 1$. Deducir que la ecuación $x^2 - 3y^2 = -1$ no tiene soluciones enteras. Encontrar una expresión para las soluciones enteras de $x^2 - 3y^2 = 1$.

Pasamos ya a abordar el problema de justificar la existencia de unidades no triviales en cuerpos cuadráticos reales arbitrarios. Notemos que en cada caso

particular esto puede justificarse encontrando una en concreto, pero el problema es razonar que siempre habrá alguna que encontrar. Para ello nos basaremos en un sencillo teorema de Dirichlet de *aproximación diofántica,* es decir, de aproximación de números irracionales por números racionales.

Es obvio que todo número irracional se puede aproximar con cualquier precisión deseada mediante un número racional. Por ejemplo, si queremos aproximar $\pi = 3.1415\ldots$ con un error menor que una milésima, basta tomar

$$r = 3.141 = \frac{3\,141}{1\,000}$$

Así, $\pi - r < 0.0006$. La aproximación es buena en el sentido de que el error es pequeño (incluso en términos relativos, el error es menor que un 2%), pero es mala en el sentido de que es fácil aproximar un número irracional con error menor que una milésima mediante un número racional de denominador 1 000. En general, una aproximación racional de un número irracional α que cumpla

$$\left|\alpha - \frac{u}{v}\right| < \frac{1}{v}$$

no tiene nada de extraordinario. El teorema de Dirichlet garantiza la existencia de aproximaciones diofánticas arbitrariamente buenas en el sentido de que mejoran arbitrariamente esta última desigualdad:

Teorema 9.9 (Dirichlet) *Si α es un número irracional y M es un número natural no nulo, existen enteros u y v con $0 < v < M$ de modo que*

$$\left|\alpha - \frac{u}{v}\right| \leq \frac{1}{Mv}.$$

La demostración proporciona un método para encontrar r, así que vamos a aplicarlo al caso de $\alpha = \pi$ y $M = 10$, pero veremos que el proceso vale en cualquier caso. La tabla siguiente contiene todos los cálculos necesarios:

	π	2π	3π	4π	5π	6π	7π	8π	9π	
	3.14159	6.28319	9.42478	12.5664	15.708	18.8496	21.9911	25.1327	28.2743	
0	$8\pi - 25$	$\pi - 3$	$9\pi - 28$	$2\pi - 6$	$3\pi - 9$	$4\pi - 12$	$5\pi - 15$	$6\pi - 18$	$7\pi - 21$	1
0	0.13274	0.14159	0.27433	0.28318	0.42477	0.56637	0.70796	0.84955	0.99114	1
	$8\pi - 25$	$22 - 7\pi$	$8\pi - 25$	$22 - 7\pi$	$\pi - 3$	$\pi - 3$	$\pi - 3$	$\pi - 3$	$\pi - 3$	$22 - 7\pi$
	0.13	0.008	0.13	0.008	0.14	0.14	0.14	0.14	0.14	0.008

Consideramos los números $\alpha, 2\alpha, \ldots, (M-1)\alpha$, les restamos su parte entera y ordenamos los resultados de menor a mayor. Si consideramos también 0 y 1, tenemos $N + 1$ números entre 0 y 1, necesariamente distintos, pues si dos de ellos fueran iguales, despejando α obtendríamos que es un número racional.

Por lo tanto, la diferencia entre dos de estos números consecutivos tiene que ser $\leq 1/M$ el algún caso, pues no es posible dividir el intervalo comprendido entre 0 y 1 es M intervalos de longitud $> 1/M$.

Elegimos, pues, dos números consecutivos cuya distancia sea $\leq 1/M$ (para obtener la mejor aproximación posible, tomamos, de hecho, la menor de estas

distancias). La última fila de la tabla contiene la distancia entre cada número y su precedente, y vemos que el mínimo se alcanza en $22 - 7\pi = 0.008\ldots$, que es mucho menor que $1/10$. Esto equivale a que

$$\frac{22}{7} - \pi < 0.008\ldots,$$

luego

$$r = \frac{22}{7} = 3.1428\ldots$$

es la aproximación racional buscada. Notemos que el 7 se obtiene de restar dos números distintos entre 0 y $M-1$, por lo que siempre se cumple que $0 < v < M$. Según indicábamos, es claro que este procedimiento puede aplicarse a cualquier número α con cualquier número M. ∎

En términos absolutos, la aproximación que hemos obtenido no es muy buena, pero aproxima π con un error menor que una centésima y el denominador es menor que 10. És en ese sentido en el que es buena. Si repetimos el proceso con $M = 110$ obtenemos

$$\pi \approx \frac{333}{106} \approx 3.14151,$$

que da lugar a un error menor que una diezmilésima con un denominador del orden de 100. En cambio, si empleamos $M = 120$ obtenemos la notable aproximación

$$\pi \approx \frac{355}{113} \approx 3.141592921$$

que da lugar a un error menor que una millonésima con un denominador del orden de 100.

Milü o el número de Metius El matemático y astrónomo chino Zu Chongzhi (429–500) determinó que

$$3.1415926 < \pi < 3.1415927$$

y encontró las aproximaciones racionales $22/7$ y $355/113$, a las que llamó *Yuelü* (aproximación basta) y *Milü* (aproximación precisa). Para ello Zu tuvo que aproximar el círculo con polígonos regulares de $2^{13} \cdot 3$ lados. En occidente la aproximación $355/113$ no fue conocida hasta 1585, cuando la descubrió el matemático neerlandés Adriaan Anthonisz, aunque el resultado no se conoció hasta que lo publicó su hijo, y por eso $355/113$ fue conocido como el *número de Metius* (pues Metius era el sobrenombre con que se conocía al hijo). ∎

Consideremos ahora un cuerpo cuadrático real $\mathbb{Q}(\sqrt{d}\,)$. Por el teorema anterior, fijado un número natural $M \geq 1$, existen números naturales u y v tales que $0 < v < M$ y

$$\left|\frac{u}{v} - \sqrt{d}\right| \leq \frac{1}{Mv}.$$

Equivalentemente, $|u - v\sqrt{d}| \leq 1/M$. Si llamamos $\alpha = u - v\sqrt{d}$, entonces

$$|\bar{\alpha}| = |\alpha + 2v\sqrt{d}| \leq \frac{1}{M} + 2M\sqrt{D} < 3M\sqrt{d},$$

luego

$$|\,\mathrm{N}(\alpha)| = |\alpha\bar{\alpha}| < \frac{1}{M}3M\sqrt{d} = 3\sqrt{d}.$$

Pero $\mathrm{N}(\alpha)$ es un entero racional, luego sólo puede tomar un número finito de valores.

Fijado, por ejemplo, $M_0 = 1$, podemos encontrar $\alpha_0 = u_0 - v_0\sqrt{d}$ que cumpla lo requerido, pero tomando entonces M_1 tal que $|\alpha_0| > 1/M_1$, un entero α_1 que cumple lo requerido con M_1 tendrá que ser distinto del anterior, y tomando un M_2 suficientemente grande, podemos forzar a que un entero α_2 que cumpla lo requerido para M_2 sea distinto de los dos anteriores. Así vamos obteniendo una sucesión enteros algebraicos $\alpha_n = u_n - v_n\sqrt{d}$ distintos dos a dos. Como $\mathrm{N}(\alpha_n)$ sólo puede tomar un número finito de valores, tiene que haber infinitos para los que la norma tome un mismo valor N. Como los pares $(\bar{u}_n, \bar{v}_n)$ de clases de restos módulo N sólo pueden tomar N^2 valores, tiene que haber infinitos valores de n para los que el par de clases de restos sea el mismo. Nos basta con tomar dos valores distintos $\alpha = u - v\sqrt{d}$ y $\alpha' = u' - v'\sqrt{d}$ tales que

$$\mathrm{N}(\alpha) = \mathrm{N}(\alpha'), \quad u \equiv u' \pmod{N}, \quad v \equiv v' \pmod{N}.$$

Tomamos

$$\epsilon = \frac{\alpha}{\alpha'} = \frac{u - v\sqrt{d}}{u' - v'\sqrt{d}} = \frac{(u - v\sqrt{d})(u' - v'\sqrt{d}))}{N} = \frac{uu' + vv'd}{N} - \frac{(uv' + vu')}{N}\sqrt{d}$$

Las congruencias muestran que ϵ está en $\mathbb{Z}\big[\sqrt{d}\,\big]$, luego es un entero algebraico, y claramente $\mathrm{N}(\epsilon) = 1$, luego es una unidad. Además $\epsilon \neq 1$, pues $\alpha \neq \alpha'$ y $\epsilon \neq -1$, pues $\alpha \neq -\alpha'$, ya que $v, v' > 0$.

Con esto queda justificada la existencia de unidades $\epsilon \neq \pm 1$ en cualquier cuerpo cuadrático real.

Ejemplo La demostración del teorema 9.6 proporciona un procedimiento para encontrar unidades fundamentales, pero no es nada eficiente. Los ejemplos que hemos considerado se resuelven fácilmente por tanteo, pero el procedimiento que hemos encontrado para encontrar sistemáticamente unidades fundamentales es muy laborioso.

Consideremos por ejemplo el cuerpo $\mathbb{Q}\big(\sqrt{22}\,\big)$. Las unidades $\epsilon = u + v\sqrt{22}$ cumplen la ecuación $u^2 - 22v^2 = \pm 1$.

Si aplicamos el teorema de Dirichlet a $\alpha = \sqrt{22}$ con $M = 1$ obtenemos el entero $\alpha_0 = -5 + \sqrt{22}$, que cumple $\mathrm{N}(\alpha_0) = 3$ y $1/|\alpha_0| = 3.23$, luego para encontrar otro distinto tenemos que aplicar el teorema de Dirichlet con $M = 4$. El resultado es $\alpha_1 = -14 + 3\sqrt{22}$, que cumple $\mathrm{N}(\alpha_1) = -2$. Necesitamos encontrar dos enteros de la misma norma (en realidad nos bastaría que el valor absoluto de la norma fuera el mismo), así que tenemos que continuar. Como

$1/|\alpha_1| = 14.03$, para encontrar un entero distinto necesitamos $M = 15$. Al aplicar el teorema de Dirichlet con este valor resulta $\alpha_2 = 61 - 13\sqrt{22}$, que tiene norma $\mathrm{N}(\alpha_2) = 3$. Vemos entonces que nos sirven

$$\alpha_0 = -5 + \sqrt{22}, \qquad \alpha_2 = 61 - 13\sqrt{22},$$

pues se cumplen también las congruencias

$$-5 \equiv 61 \text{ (mód 3)}, \qquad 1 \equiv 13 \text{ (mód 3)}.$$

De no haber sido así, habríamos tenido que continuar, pero esto basta para asegurar que el cociente

$$\epsilon = \frac{\bar{\alpha}_2}{\alpha_1} = \frac{61 + 13\sqrt{22}}{-5 + \sqrt{22}} = -197 - 42\sqrt{22}$$

es entero y, de hecho, una unidad. Como es negativa, pasamos a

$$\eta = 197 + 42\sqrt{22}.$$

Así hemos encontrado una unidad no trivial $\eta > 1$ tal que $\mathrm{N}(\eta) = 1$. Sucede que es la unidad fundamental, pero para comprobarlo no conocemos mejor método que comprobar que no existe ninguna unidad $u + v\sqrt{22}$ con $0 < u < 197$ y $0 < v < 42$. Por ejemplo, si intentamos con $v = 41$ vemos que $x^2 - 22 \cdot 41^2 = \pm 1$ no tiene soluciones enteras, y del mismo modo se descartan las otras 40 posibilidades.

Un ordenador puede hacer todas las comprobaciones sin dificultad, pero acaba mucho antes si lo programamos para ir dando valores a $u^2 - 22v^2$ y que se detenga cuando obtenga una solución distinta de $(1, 0)$ en la que la expresión tome el valor ± 1. En el capítulo siguiente veremos un método mucho más sencillo para encontrar unidades fundamentales usando resultados de aproximación diofántica más potentes que el teorema de Dirichlet. ∎

9.3 Fallos en la factorización única

Hemos visto que los anillos de enteros de los cuerpos cuadráticos cumplen la condición necesaria dada por el teorema 2.28 para que puedan tener factorización única. Sin embargo, dicha condición dista mucho de ser suficiente. La tabla 9.1 contiene ejemplos de factorizaciones no únicas en cuerpos cuadráticos imaginarios y muestra que son relativamente frecuentes.

Ejemplo Analicemos con detalle el caso de

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

Notemos que no es inmediato que sea un ejemplo de factorización no única. En principio, podría ser un caso análogo a

$$10 = 2 \cdot 5 = (1 + 3i)(1 - 3i),$$

Tabla 9.1: Factorizaciones no únicas en cuerpos cuadráticos imaginarios

d					
-5	6	$=$	$2\cdot 3$	$=$	$\big(1+\sqrt{-5}\,\big)\big(1-\sqrt{-5}\,\big)$
-6	6	$=$	$2\cdot 3$	$=$	$-\sqrt{-6}^{\,2}$
-10	14	$=$	$2\cdot 7$	$=$	$\big(2+\sqrt{-10}\,\big)\big(2-\sqrt{-10}\,\big)$
-13	14	$=$	$2\cdot 7$	$=$	$\big(1+\sqrt{-13}\,\big)\big(1-\sqrt{-13}\,\big)$
-14	15	$=$	$3\cdot 5$	$=$	$\big(1+\sqrt{-14}\,\big)\big(1-\sqrt{-14}\,\big)$
-15	4	$=$	$2\cdot 2$	$=$	$\left(\frac{1+\sqrt{-15}}{2}\right)\left(\frac{1-\sqrt{-15}}{2}\right)$
-17	18	$=$	$2\cdot 3\cdot 3$	$=$	$\big(1+\sqrt{-17}\,\big)\big(1-\sqrt{-17}\,\big)$
-21	22	$=$	$2\cdot 11$	$=$	$\big(1+\sqrt{-21}\,\big)\big(1-\sqrt{-21}\,\big)$
-22	26	$=$	$2\cdot 13$	$=$	$\big(1+\sqrt{-22}\,\big)\big(1-\sqrt{-22}\,\big)$
-23	6	$=$	$2\cdot 3$	$=$	$\left(\frac{1+\sqrt{-23}}{2}\right)\left(\frac{1-\sqrt{-23}}{2}\right)$
-26	27	$=$	$3\cdot 3\cdot 3$	$=$	$\big(1+\sqrt{-26}\,\big)\big(1-\sqrt{-26}\,\big)$
-29	30	$=$	$2\cdot 3\cdot 5$	$=$	$\big(1+\sqrt{-29}\,\big)\big(1-\sqrt{-29}\,\big)$
-30	34	$=$	$2\cdot 17$	$=$	$\big(2+\sqrt{-30}\,\big)\big(2-\sqrt{-30}\,\big)$

que no contradice la factorización única de $\mathbb{Z}[i]$. Lo que sucede en este caso es que los factores no son irreducibles, sino que se descomponen en irreducibles (primos) del modo siguiente:

$$10=\overbrace{(1+i)(1-i)}^{2}\overbrace{(2+i)(2-i)}^{3}=\overbrace{(1+i)(2+i)}^{1+3i}\overbrace{(1-i)(2-i)}^{1-3i},$$

de modo que en ambos casos tenemos la misma descomposición en factores primos para el 10.

No sucede lo mismo con las factorizaciones del 6 en $\mathbb{Q}\big(\sqrt{-5}\,\big)$. Observemos que

$$\mathrm{N}(2)=4,\quad \mathrm{N}(3)=9,\quad \mathrm{N}(1+\sqrt{-5})=6,\quad \mathrm{N}(1-\sqrt{-5})=6.$$

Si estos números no fueran irreducibles, se descompondrían en producto de dos irreducibles de norma 2 en el caso del 2, de norma 3 en el caso del 3 y uno de norma 2 y otro de norma 3 en el caso de los dos últimos números. Pero

$$\mathrm{N}(a+b\sqrt{-5})=a^2+5b^2,$$

y es claro que no hay enteros de norma 2 ni de norma 3.

Así pues, estamos realmente ante dos descomposiciones de 6 en factores irreducibles no asociados (pues las únicas unidades son ± 1). Vemos, de hecho,

Tabla 9.2: Factorizaciones no únicas en cuerpos cuadráticos reales

d					
10	6	$=$	$2\cdot 3$	$=$	$\big(4+\sqrt{10}\,\big)\big(4-\sqrt{10}\,\big)$
15	10	$=$	$2\cdot 5$	$=$	$\big(5+\sqrt{15}\,\big)\big(5-\sqrt{15}\,\big)$
26	10	$=$	$2\cdot 5$	$=$	$\big(6+\sqrt{26}\,\big)\big(6-\sqrt{26}\,\big)$
30	6	$=$	$2\cdot 3$	$=$	$\big(6+\sqrt{30}\,\big)\big(6-\sqrt{30}\,\big)$

que los cuatro factores son irreducibles, pero no primos, pues cada uno de ellos divide al miembro opuesto y no divide a ninguno de los dos factores en que se descompone. ∎

Ejercicio: Comprobar que las otras factorizaciones de la tabla 9.1 son realmente ejemplos de factorizaciones no únicas.

En los cuerpos cuadráticos reales el trabajo se complica debido a que las normas pueden ser negativas, lo que impide descartar tan fácilmente casos como $\mathrm{N}(x)=2$ en $\mathbb{Q}\big(\sqrt{-5}\,\big)$. Pese a ello, la tabla 9.2 contiene algunos ejemplos que podemos comprobar.

Ejemplo En $\mathbb{Q}\big(\sqrt{10}\,\big)$ tenemos que

$$\mathrm{N}(2)=4,\quad \mathrm{N}(3)=9,\quad \mathrm{N}\big(4+\sqrt{10}\,\big)=\mathrm{N}\big(4-\sqrt{10}\,\big)=6.$$

Vamos a ver que no hay enteros de norma igual a ± 2 ni a ± 3, por lo que los cuatro serán irreducibles y no asociados (pues no tienen la misma norma).

Si existiera un entero $x=a+b\sqrt{10}$ tal que

$$\mathrm{N}(a+b\sqrt{10})=a^2-10b^2=\pm 2,\pm 3$$

entonces tendría que ser $a^2\equiv \pm 2$ o ± 3 (mód 10), o lo que es lo mismo,

$$a^2\equiv 2,3,7,8 \text{ (mód 10)}.$$

Sin embargo, los cuadrados módulo 10 son: $0,1,4,9,6,5,6,9,4,1$, luego la congruencia es imposible. Los otros casos se prueban de modo similar. ∎

9.4 Cuerpos cuadráticos euclídeos

A pesar de las observaciones de la sección precedente, sabemos que los anillos de enteros de los cuerpos imaginarios con $d=-1,-2,-3$ son dominios euclídeos, y vamos a ver que no son los únicos. Concretamente, sabemos que estos tres anillos son dominios euclídeos tomando como norma euclídea la norma algebraica. Por definición, una norma euclídea sólo puede tomar valores naturales, así que en el caso real no podemos considerar la norma sino, a lo sumo, el valor absoluto de la norma.

Vamos a estudiar qué debe cumplir la norma en un cuerpo cuadrático para que su valor absoluto sirva como norma euclídea. Ciertamente cumple la condición $|\,\mathrm{N}(\alpha)| \le |\,\mathrm{N}(\alpha\beta)|$, cuando α y β son enteros no nulos. Respecto a la existencia de división euclídea, la situación es la siguiente:

> *Si $d \neq 1$ es libre de cuadrados, el anillo de enteros de $\mathbb{Q}\big(\sqrt{d}\,\big)$ es un dominio euclídeo con el valor absoluto de la norma si y sólo si para todo α en $\mathbb{Q}\big(\sqrt{d}\,\big)$ existe un entero γ tal que $|\,\mathrm{N}(\alpha-\gamma)| < 1$.*

En efecto, si se cumple esta condición, dados dos enteros Δ y δ con $\delta \neq 0$, la aplicamos a $\alpha = \Delta/\delta$ y llamamos $\epsilon = (\alpha-\gamma)\delta = \Delta - \delta\gamma$, de modo que $\Delta = \delta\gamma + \epsilon$ y $|\,\mathrm{N}(\epsilon)| = |\,\mathrm{N}(\alpha-\gamma)||\,\mathrm{N}(\delta)| < \mathrm{N}(\delta)$.

Recíprocamente, si el anillo de enteros es un dominio euclídeo, todo γ en $\mathbb{Q}\big(\sqrt{d}\,\big)$ puede expresarse como $\alpha = \Delta/\delta$, donde Δ es entero y δ es un número natural. Efectuando la división euclídea $\Delta = \delta\gamma + \epsilon$, con $|\,\mathrm{N}(\epsilon)| < |\,\mathrm{N}(\delta)|$, tenemos que $|\,\mathrm{N}(\alpha-\gamma)| = |\,\mathrm{N}(\epsilon/\delta)| < 1$.

Para estudiar si se cumple esta propiedad, podemos expresar $\alpha = r + s\sqrt{d}$, donde r y s son números racionales, y tenemos que encontrar un $\gamma = x + y\sqrt{d}$, donde x, y son enteros racionales, o bien (sólo en el caso $d \equiv 1 \pmod 4$) ambos son semienteros no enteros, de modo que

$$|(r-x)^2 + d(s-y)^2| < 1.$$

Ahora bien, si expresamos $r = m + r'$, $s = n + s'$, de modo que m y n son enteros racionales y $|r'| \le 1/2$, $|s'| \le 1/2$, vemos que si podemos encontrar x' e y' en las condiciones requeridas para r' y s', entonces $m + x'$ y $n + y'$ las cumplen para r y s, luego no perdemos generalidad si suponemos que $|r| \le 1/2$ y $|s| \le 1/2$.

Más aún, si $-1/2 \le r < 0$ y encontramos x', y' para $-r$ y s, entonces $-x'$ e y' cumplen la condición para r y s, luego no perdemos generalidad si suponemos que $0 \le r \le 1/2$, e igualmente podemos suponer que $0 \le s \le 1/2$.

Si $d \equiv 1 \pmod 4$ todavía podemos decir más, pues si $1/4 < s \le 1/2$ y encontramos x' e y' para $r' = 1/2 - r$, $s' = 1/2 - s$, entonces $x = 1/2 - x'$ e $y = 1/2 - y'$ son ambos enteros o ambos semienteros y cumplen lo requerido para r y s. Por lo tanto, en este caso podemos suponer que $0 \le y \le 1/4$.

Para resumir lo que hemos obtenido definimos el rectángulo

$$F(d) = \begin{cases} \{(r,s) \in \mathbb{Q}^2 \mid 0 \le r \le 1/2,\ 0 \le s \le 1/2\} & \text{si } d \not\equiv 1 \pmod 4, \\ \{(r,s) \in \mathbb{Q}^2 \mid 0 \le r \le 1/2,\ 0 \le s \le 1/4\} & \text{si } d \equiv 1 \pmod 4, \end{cases}$$

así como el conjunto

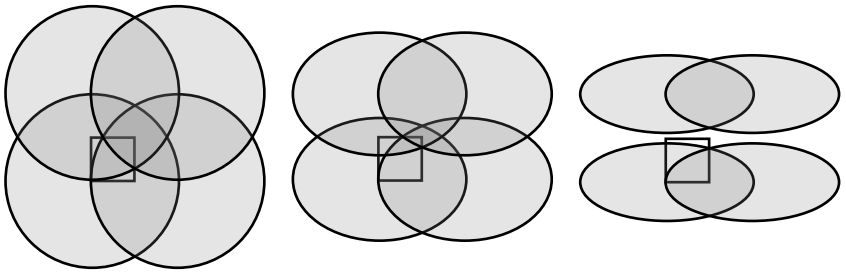
$$U(x,y) = \{(r,s) \in \mathbb{Q}^2 \mid |(r-x)^2 + d(s-y)^2| < 1\},$$

y así, lo que hemos obtenido es lo siguiente:

Si $d \neq 1$ es libre de cuadrados, el anillo de enteros de $\mathbb{Q}\big(\sqrt{d}\big)$ es un dominio euclídeo con el valor absoluto de la norma si y sólo si todo punto (r, s) del rectángulo $F(d)$ está contenido en un conjunto $U(x, y)$, donde x e y son enteros racionales o (sólo en el caso en que $d \equiv 1 \pmod 4$) ambos son semienteros no enteros.

En otras palabras, $F(d)$ tiene que poder cubrirse con conjuntos $U(x, y)$.

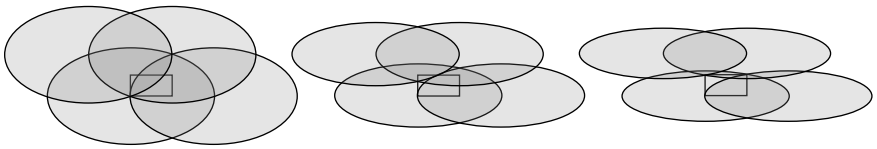
Restrinjámonos ahora al caso en que $d < 0$. Entonces $U(x, y)$ es una elipse de centro (x, y) cuyo semieje horizontal mide 1 y su semieje vertical mide $1/\sqrt{-d}$.



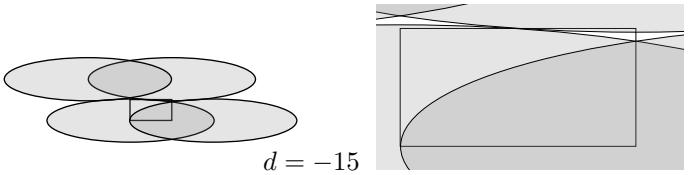
$d = -1$ $d = -2$ $d = -5$

La figura muestra el cuadrado $F(d)$ y los conjuntos $U(x, y)$ con centros cercanos para $d = -1$, $d = -2$ y $d = -5$. En los dos primeros casos vemos que $U(0, 0)$ basta para cubrir todo el cuadrado, mientras que en el caso $d = -5$ vemos que una parte del cuadrado queda sin cubrir (y cualquier otro conjunto $U(x, y)$ no representado ni siquiera toca al cuadrado). Por lo tanto, el anillo de enteros de $\mathbb{Q}\big(\sqrt{-5}\big)$ no es euclídeo con el valor absoluto de la norma, ni tampoco puede serlo el de ningún otro cuerpo cuadrático imaginario con $d \not\equiv 1 \pmod 4$, ya que cada elipse correspondiente a un $d < -5$ está contenida en la correspondiente a $d = -5$ con el mismo centro, luego si éstas no cubren el cuadrado, las menores tampoco.

Ahora consideramos los valores $d \equiv 1 \pmod 4$, para los que $F(d)$ es la mitad del cuadrado que estábamos considerando, y tenemos que contemplar más centros posibles para las elipses.



$d = -3$ $d = -7$ $d = -11$



$d = -15$

Vemos que hasta $d = -11$ basta $U(0,0)$ para cubrir todo el rectángulo, mientras que para $d = -15$ una pequeña porción del rectángulo queda sin cubrir (y cualquier otra elipse válida no representada ni siquiera corta al rectángulo). Con esto hemos probado parte del teorema siguiente:

Teorema 9.10 *Sea d un número negativo libre de cuadrados. El anillo de los enteros de $\mathbb{Q}(\sqrt{d})$ es un dominio euclídeo si y sólo si d es uno de los cinco números siguientes:*

$$-1, \quad -2, \quad -3, \quad -7, \quad -11$$

En estos casos, la norma euclídea es $\phi(a) = \mathrm{N}(a)$.

DEMOSTRACIÓN: Acabamos de probar que estos cinco anillos son dominios euclídeos y que ningún otro puede serlo con la norma algebraica, pero podemos probar que tampoco pueden serlo con ninguna otra norma.

En efecto, hemos visto que los anillos de enteros para $d = -5, -6, -10$ no tienen factorización única, luego no pueden ser euclídeos. Falta descartar los valores $d < -11$ libres de cuadrados, luego, de hecho, $d \leq -13$.

Si A es el anillo de enteros de $\mathbb{Q}(\sqrt{d})$ y es un dominio euclídeo con alguna norma euclídea, podemos tomar un entero δ de norma mínima entre los enteros no nulos ni unitarios. Entonces todo entero Δ se expresa en la forma $\Delta = \delta c + r$, donde $r = 0, 1, -1$, por la elección de δ. Esto significa que el anillo de clases de restos A_δ consta únicamente de las clases $\bar{0}$, $\bar{1}$ y $-\bar{1}$.

No puede ser que $\bar{1} + \bar{1} = \bar{1}$, pues entonces $\bar{1} = \bar{0}$ y $\delta \mid 1$, luego δ sería una unidad. Si $\bar{1} + \bar{1} = \bar{0}$, entonces $\bar{2} = \bar{0}$, luego $\delta \mid 2$. En caso contrario, $\bar{1} + \bar{1} = -\bar{1}$, luego $\bar{3} = \bar{0}$ y $\delta \mid 3$.

Pero no puede ser $\delta = \pm 2, \pm 3$, ya que ω no es congruente con ± 1 o 0 módulo un entero racional $n \geq 1$ (más en general, no puede ser que $n \mid m + \omega$, pues esto significa que $m + \omega = n(a + b\omega)$ y entonces $nb = 1$, luego $n = 1$). La conclusión es que $\mathrm{N}(\delta) = 2, 3$.

Ahora bien, si $\delta = (a/2) + (b/2)\sqrt{d}$, donde a y b son enteros (ambos pares o ambos impares), tenemos que $a^2 - db^2 \leq 12$, y como $d \leq -13$, necesariamente $b = 0$ y $|a| \leq 3$, pero entonces $\delta = a/2$ es entero y no puede ser más que $\delta = 0, 1, -1$, en contra del criterio con que ha sido elegido. ∎

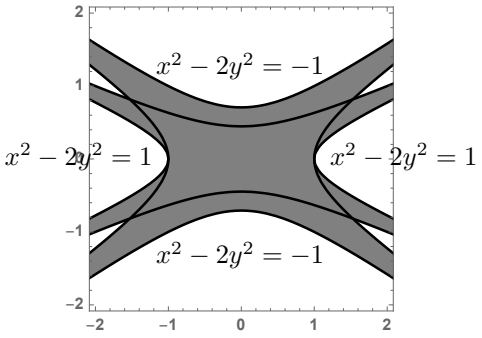
Esto no significa que los cuerpos considerados en el teorema anterior sean los únicos cuyos anillos de enteros tienen factorización única. Puede probarse que los cuerpos cuadráticos imaginarios con factorización única son exactamente los correspondientes a

$$d = -1, \quad -2, \quad -3, \quad -7, \quad -11, \quad -19, \quad -43, \quad -67 \quad -163.$$

El caso de los cuerpos cuadráticos reales es más complicado. Ahora los conjuntos $U(x, y)$ ya no son elipses, sino que están limitados por cuatro ramas parabólicas.

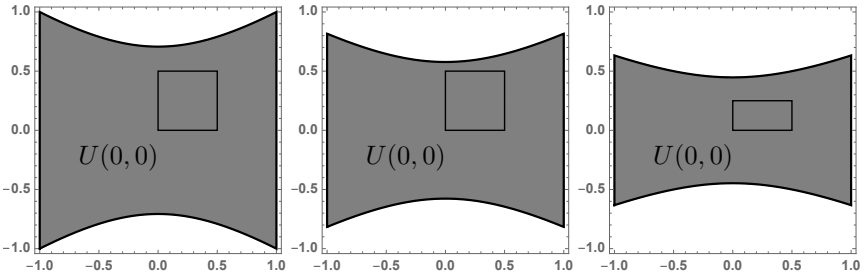
La figura siguiente muestra $U(0,0)$ para $d = 2$ y $d = 5$. Por una parte vemos que se extienden indefinidamente, por lo que para cubrir $F(d)$ no podemos limitarnos a considerar los centros cercanos, sino que pueden hacer falta centros arbitrariamente lejanos. Por otra parte, vemos que, al contrario de lo que

sucedía con las elipses, el conjunto correspondiente a $d = 5$ no está contenido en el correspondiente a $d = 2$, por lo que el hecho de que los conjuntos $U(d)$ para un cierto d no cubran a $F(d)$ no nos permite asegurar que los conjuntos correspondientes a valores mayores de d no puedan hacerlo.

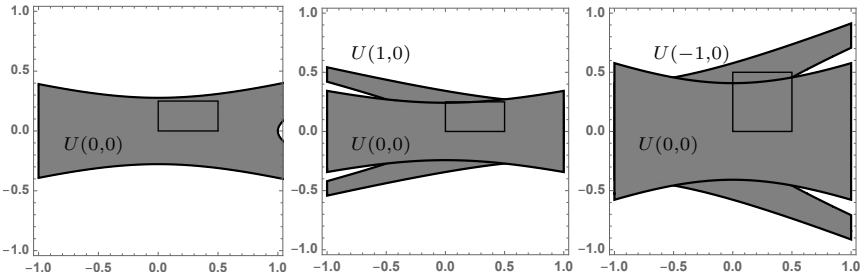


Por ejemplo, para $d = 2, 3, 5, 13$ basta $U(0, 0)$ para cubrir $F(d)$, mientras que para $d = 17$ hace falta considerar también $U(1, 0)$. Las figuras siguientes muestran los conjuntos $U(x, y)$ necesarios para cubrir $F(d)$ para distintos valores de d.

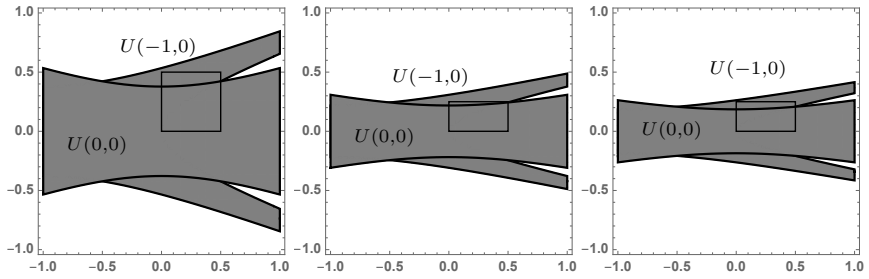
Notemos que, por ejemplo, para $d = 6$, hay un punto con $x = 1/2$ en $F(d)$ que está en la frontera de los dos conjuntos $U(0, 0)$ y $U(-1, 0)$, luego no está contenido en ninguno de los dos, pero no importa, porque podemos calcular qué punto es y resulta ser $(1/2, \sqrt{5/24})$, que es irracional, luego no necesitamos cubrirlo. Lo mismo sucede en los casos $d = 7, 21, 29$, así como en $d = 11$.



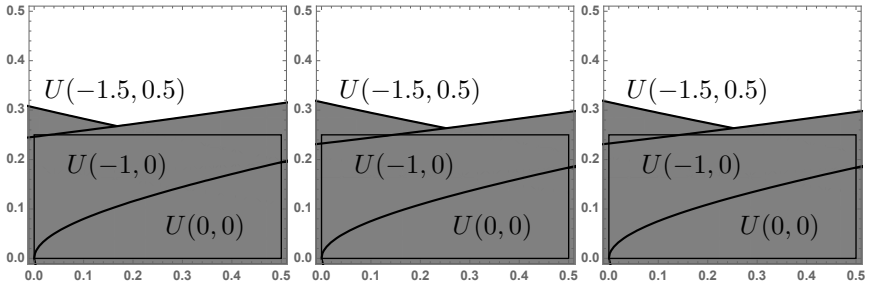
$d = 2$ $\quad$ $d = 3$ $\quad$ $d = 5$



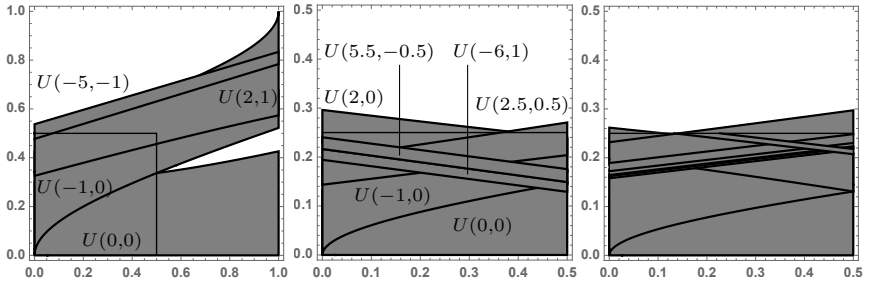
$d = 13$ $\quad$ $d = 17$ $\quad$ $d = 6$



$d = 7$ $d = 21$ $d = 29$



$d = 33$ $d = 37$ $d = 41$



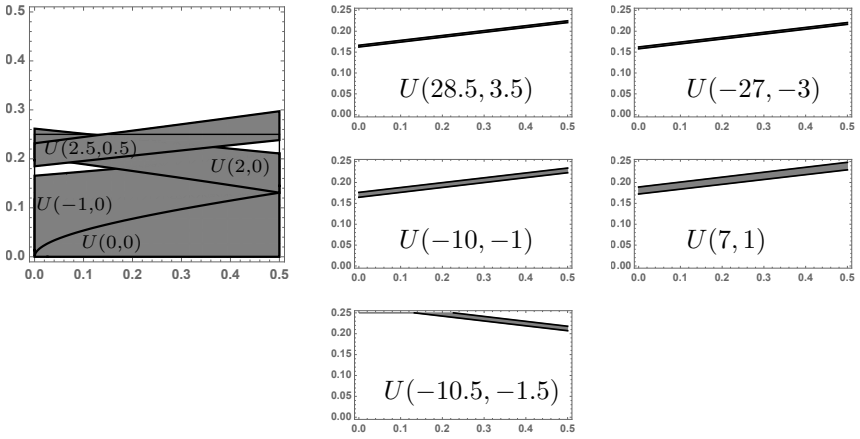
$d = 11$ $d = 57$ $d = 73$

Los casos $d = 57, 73$ son demasiado complejos como para que la mera inspección de la imagen resulte concluyente, y es necesario examinar algebraicamente los distintos conjuntos para asegurar que realmente cubren todo $F(d)$ y, en particular, que ningún punto de $F(d)$ queda únicamente en las fronteras de los conjuntos $U(x, y)$ considerados. Por ejemplo, para $d = 73$ necesitamos considerar los conjuntos

$$U(0,0), \quad U(-1,0), \quad U(2,0), \quad U(2.5,0.5),$$

$$U(7,1), \quad U(-10,-1), \quad U(-27,-3), \quad U(28.5,3.5), \quad U(-10,5,-1.5).$$

Como muestra la figura siguiente, los cuatro primeros cubren la mayor parte de $F(73)$, pero para cubrir las dos porciones restantes es necesario recurrir a conjuntos con centros bastante alejados.



La figura siguiente muestra la superposición de $U(28.5, 3.5)$ y $U(-27, -3)$.

Dentro del rectángulo $F(73)$, el conjunto $U(28.5, 3.5)$ es una banda limitada por las gráficas de las funciones

$$3.5 - \sqrt{\frac{(x-28.5)^2+1}{73}}, \qquad 3.5 - \sqrt{\frac{(x-28.5)^2-1}{73}},$$

mientras que $U(-27, -3)$ está limitado por las gráficas de las funciones

$$-3 + \sqrt{\frac{(x+27)^2-1}{73}}, \qquad -3 + \sqrt{\frac{(x+27)^2+1}{73}}.$$

En $x = 0$, los valores que toman son $[0.162275, 0.166381]$ en el primer caso y $[0.157943, 0.162278]$ en el segundo. Así pues, vemos que las dos bandas empiezan solapándose, y si igualamos

$$\sqrt{\frac{(x+27)^2+1}{73}} = 3.5 - \sqrt{\frac{(x-28.5)^2+1}{73}},$$

la solución es $x = 3/4$, que está fuera de $F(d)$, lo cual significa que las dos bandas no dejan de solaparse mientras están sobre $F(d)$, luego su unión es una banda más amplia limitada por las funciones

$$-3 + \sqrt{\frac{(x+27)^2-1}{73}}, \qquad 3.5 - \sqrt{\frac{(x-28.5)^2-1}{73}}.$$

Un análisis análogo nos da que $U(28.5, 3.5)$ se solapa con $U(-10, -1)$, que es la banda limitada por las funciones

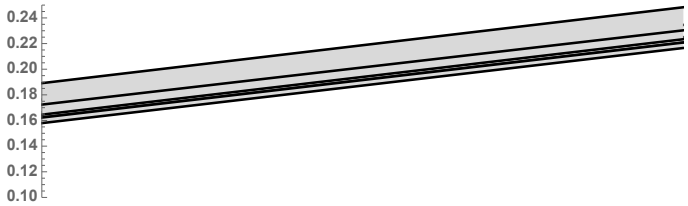
$$-1 + \sqrt{\frac{(x+10)^2 - 1}{73}}, \qquad -1 + \sqrt{\frac{(x+10)^2 + 1}{73}},$$

y a su vez, esta banda se solapa con $U(7, 1)$, que es la banda limitada por

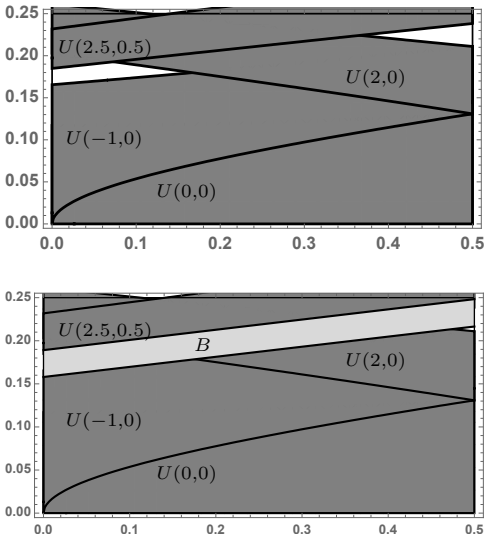
$$1 - \sqrt{\frac{(x-7)^2 + 1}{73}}, \qquad 1 - \sqrt{\frac{(x-7)^2 - 1}{73}}.$$

La figura siguiente muestra la superposición B de las cuatro bandas, que está limitada por las funciones

$$-3 + \sqrt{\frac{(x+27)^2 - 1}{73}}, \qquad 1 - \sqrt{\frac{(x-7)^2 - 1}{73}}.$$



A su vez, comprobaciones análogas muestran que la frontera superior de B se solapa con la frontera inferior de $U(2.5, 0.5)$ en todo $F(73)$ y que su frontera inferior se solapa con la frontera superior de $U(-1, 0)$ hasta $x = 0.255$, que es posterior a $x = 0.166$, que es la "punta" derecha del agujero izquierdo que debemos cubrir. Esto prueba que B cubre por completo dicho agujero izquierdo.

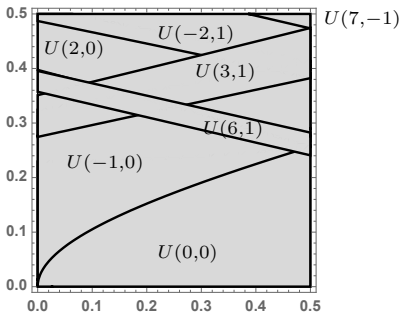


Sin embargo, la frontera inferior de B corta a la frontera superior de $U(2,0)$ en $x = 0.474$, y esto hace que deje sin cubrir un pequeño triángulo del "agujero" derecho, como muestra la figura. Comprobaciones análogas muestran que el último conjunto, $U(-10.5,-1.5)$, cubre completamente este último hueco, y esto completa la prueba de que el anillo de enteros de $\mathbb{Q}\big(\sqrt{73}\big)$ es un dominio euclídeo.

Similarmente puede probarse que el anillo de enteros de $\mathbb{Q}\big(\sqrt{19}\big)$ es un dominio euclídeo. La figura siguiente muestra siete conjuntos que casi cubren la totalidad de $F(19)$, pero en realidad queda sin cubrir un triángulo diminuto a la izquierda, entre $U(3,1)$ y $U(6,1)$. Para cubrirlo es necesario recurrir a $U(19,-4)$, $U(-90,21)$ y $U(-430,99)$. Estos tres conjuntos forman tres bandas que se solapan entre sí para formar una banda más amplia comprendida entre

$$99 - \sqrt{\frac{(x+430)^2+1}{19}} \qquad \text{y} \qquad -4 + \sqrt{\frac{(x-19)^2+1}{19}},$$

la cual se solapa a su vez con $U(6,1)$ y basta para cubrir el pequeño agujero.



Con esto hemos demostrado una parte del teorema siguiente:

Teorema 9.11 *Los únicos cuerpos cuadráticos reales $\mathbb{Q}\big(\sqrt{d}\big)$ cuyos anillos de enteros son dominios euclídeos con el valor absoluto de la norma algebraica son los correspondientes a los valores siguientes de d:*

2, 3, 5, 6, 7, 11, 13, 17, 19, 21, 29, 33, 37, 38, 41, 57, 73.

No estamos en condiciones de demostrar que son los únicos, pero hemos probado que todos ellos son ciertamente dominios euclídeos. Por otra parte, puede probarse que existen cuerpos cuadráticos reales que son euclídeos con una norma diferente de la algebraica. Dos ejemplos son los correspondientes a $d = 14$ y $d = 69$.

9.5 Factorización única en cuerpos cuadráticos

Ahora enunciamos una versión general del teorema 4.6 y los resultados análogos que hemos probado en algunos casos particulares.

Supongamos que $k = \mathbb{Q}(\sqrt{d}\,)$ es un cuerpo cuadrático y que $\mathbb{Z}[\omega]$ es su anillo de enteros. No necesitamos suponer que $\omega = \sqrt{d}$ o que $\omega = (1+\sqrt{d})/2$. Basta con que todo entero de k se exprese de forma única como $a + b\omega$, para ciertos enteros racionales a y b. Tampoco necesitamos suponer que $\mathbb{Z}[\omega]$ sea un dominio euclídeo, sino que basta con que sea un dominio de factorización única.

Si el lector ha asimilado los argumentos que hemos dado en los distintos casos particulares de esta situación que hemos discutido, debería poder justificar por sí mismo las afirmaciones siguientes:

1. *Todo primo cuadrático π (es decir, todo primo de $\mathbb{Z}[\omega]$) divide a un único primo racional p, de modo que $|\,\mathrm{N}(\pi)| = p$ o $|\,\mathrm{N}(\pi)| = p^2$.*

2. *Si p es un primo racional, o bien es un primo cuadrático, o bien factoriza como $p = \pi_1\pi_2$, donde π_1 y π_2 son primos cuadráticos.*

   En el primer caso se dice que p *se conserva primo* en $\mathbb{Z}[\omega]$ (o, por abuso de lenguaje, en $\mathbb{Q}(\sqrt{d}\,)$), mientras que en el segundo se dice que p se *ramifica* si π_1 y π_2 son asociados (y entonces $p = \epsilon\pi^2$, para cierta unidad ϵ), o que *se escinde,* si π_1 y π_2 no son asociados.

3. *Si $f(\omega)$ es el polinomio mínimo de ω (que tiene coeficientes enteros, porque ω es un entero algebraico), p es un primo racional y $\bar{c}$ es una raíz de $\bar{f}(x)$ en $\mathbb{Z}_p[x]$, entonces $\pi = (p, \omega - c)$ es un primo cuadrático, el único que cumple $\pi \mid p$ y $\omega \equiv c$ (mód π).*

   Como esto es un hecho fundamental recordamos aquí la prueba, aunque el lector haría bien en tratar de obtenerla por sí mismo, comparando si es preciso con la prueba de 4.6: en $\mathbb{Z}_p[x]$ tenemos que

   $$\bar{f}(x) = (x - \bar{c})(x - \bar{c}\,'),$$

   para cierto entero racional c', lo que equivale a que

   $$f(x) = (x - c)(x - c') + pq(x),$$

   para cierto polinomio $q(x)$. Evaluando en ω queda que

   $$(\omega - c)(\omega - c') = -pq(\omega),$$

   luego $p \mid (\omega - c)(\omega - c')$. Pero p no puede dividir a ninguno de los factores. Por ejemplo, si $p \mid \omega - c$, tenemos que $\omega - c = p(a + b\omega)$, luego, por la unicidad de la representación, $pb = 1$, lo cual es absurdo.

   Esto significa que p no es primo, sino que se descompone como $p = \pi\pi'$, para ciertos primos cuadráticos π y π', asociados o no. Pero no pueden dividir ambos al mismo factor, ya que entonces éste sería divisible entre p, por lo que tiene que ser $\pi \mid \omega - c$ y $\pi' = \omega - c'$. Así pues, $\pi = (p, \omega - c)$ y $\omega \equiv c$ (mód π). La unicidad se prueba trivialmente.

4. *Si π es primo en $\mathbb{Z}[\omega]$, entonces el anillo de clases de restos $k = \mathbb{Z}[\omega]_\pi$ es un cuerpo.*

   No podemos aplicar el teorema 4.1 porque no estamos suponiendo que $\mathbb{Z}[\omega]$ sea un dominio euclídeo, pero hay otro argumento sencillo alternativo. La definición de primo implica inmediatamente que k es un dominio íntegro (si $\bar\alpha\bar\beta = 0$, entonces $\pi \mid \alpha\beta$, luego $\pi \mid \alpha$ o $\pi \mid \beta$, luego $\bar\alpha = 0$ o $\bar\beta = 0$).

   Por otro lado, si p es el primo racional que cumple $\pi \mid p$, entonces k tiene a lo sumo p^2 elementos, ya que todo elemento de k se puede expresar como $\bar a + \bar b\bar\omega$, con $0 \le a, b < p$. Ahora basta aplicar el teorema siguiente:

Teorema 9.12 *Todo dominio íntegro finito es un cuerpo.*

Demostración: Sea A un dominio íntegro finito. Sólo tenemos que probar que todo el elemento a de A no nulo tiene inverso. Ahora bien, si $a_1, \ldots, a_n$ son todos los elementos de A, sin repeticiones, entonces los elementos $aa_1, \ldots, aa_n$ son distintos dos a dos, ya que si $aa_i = aa_j$, entonces $a(a_i - a_j) = 0$, luego $a_i - a_j = 0$, luego $i = j$. Por lo tanto, son todos los elementos de A, luego uno de ellos es $aa_i = 1$, luego a tiene inverso. ∎

Con esto ya es fácil probar:

Teorema 9.13 *Sea $\mathbb{Q}\big(\sqrt{d}\,\big)$ un cuerpo cuadrático y $\mathbb{Z}[\omega]$ su anillo de enteros algebraicos. Si éste es un dominio de factorización única, $f(x)$ es el polinomio mínimo de ω y p es un primo racional, entonces:*

1. *Si $\bar f(x) = (x - \bar c)^2$ en $\mathbb{Z}_p[x]$, entonces $p = \epsilon\pi^2$, donde $\pi = (p, \omega - c)$ es un primo cuadrático que cumple $\omega \equiv c \pmod\pi$.*
2. *Si $\bar f(x) = (x - \bar c_1)(x - \bar c_2)$ en $\mathbb{Z}_p[x]$, donde $c_1 \not\equiv c_2 \pmod p$, entonces $p = \pi_1\pi_2$, donde $\pi_i = (p, \omega - c_i)$ son primos cuadráticos no asociados tales que $\omega \equiv c_j \pmod{\pi_j}$.*
3. *Si $\bar f(x)$ es irreducible en $\mathbb{Z}_p[x]$, entonces p se conserva primo y ω no es congruente módulo p con ningún entero racional.*

Ejercicio: Probar que, en las condiciones del teorema anterior, si π es un primo cuadrático, el cuerpo de clases de restos módulo π tiene $|\,\mathrm{N}(\pi)|$ elementos.

Ejemplo Vamos a estudiar la factorización de los primos racionales en el anillo $\mathbb{Z}\big[\sqrt{6}\,\big]$. Para $p = 2, 3$ tenemos que $x^2 - 6 \equiv x^2 \pmod p$, por lo que 2 y 3 se ramifican. Para $p > 3$ tenemos que Δ_p es no nulo módulo p, luego el polinomio $x^2 - 6$ tiene dos raíces distintas módulo 6 o no tiene ninguna. Concretamente, p se escinde si y sólo si $(6/p) = 1$. Ahora podemos usar la ley de reciprocidad cuadrática a través de la sexta propiedad del símbolo de Jacobi que probamos tras la definición 7.6. Esta nos asegura que $(6/p)$ depende únicamente del resto de p módulo 24. Esto nos permite construir la tabla siguiente:

$p \pmod{24}$	-11	-7	-5	-1	1	5	7	11
$(6/p)$	-1	-1	1	1	1	1	-1	-1

Por ejemplo, si $p \equiv 1 \pmod{24}$, tenemos que

$$\left(\frac{6}{p}\right) = \left(\frac{6}{1}\right) = 1.$$

Si $p \equiv -1 \pmod{24}$, entonces, como el "denominador" del símbolo de Jacobi tiene que ser impar y positivo, hacemos:

$$\left(\frac{6}{p}\right) = \left(\frac{6}{23}\right) = \left(\frac{2}{23}\right)\left(\frac{3}{23}\right) = -\left(\frac{23}{3}\right) = -\left(\frac{2}{3}\right) = 1.$$

Igualmente se calculan los demás valores de la tabla. Veamos un ejemplo más: si $p \equiv -7 \pmod{24}$, entonces

$$\left(\frac{6}{p}\right) = \left(\frac{6}{17}\right) = \left(\frac{2}{17}\right)\left(\frac{3}{17}\right) = \left(\frac{17}{3}\right) = \left(\frac{2}{3}\right) = -1$$

Así pues, los primos racionales que se escinden en $\mathbb{Z}\big[\sqrt{6}\,\big]$ son los que cumplen $p \equiv \pm 1, \pm 5 \pmod{24}$. ∎

Ejercicio: ¿Podemos afirmar que un primo p es de la forma $p = x^2 - 6y^2$ si y sólo si $p = 2, 3$ o $p \equiv \pm 1, \pm 5 \pmod{24}$?

Ejercicio: Estudiar la factorización de primos en otros anillos de enteros algebraicos con factorización única, reales e imaginarios. Tratar de formular conjeturas sobre las clases de restos corresponden a primos que se escinden y las que corresponden a primos que se conservan.

El carácter de un cuerpo cuadrático Vamos a introducir algunos conceptos que nos permitirán sistematizar el uso de la ley de reciprocidad cuadrática para determinar el carácter de cada primo racional en un cuerpo cuadrático (es decir, si se ramifica, se escinde o se conserva).

En primer lugar definimos el *discriminante* de un cuerpo $k = \mathbb{Q}\big(\sqrt{d}\,\big)$ como[1]

$$\Delta_k = \begin{cases} d & \text{si } d \equiv 1 \pmod{4}, \\ 4d & \text{si } d \not\equiv 1 \pmod{4}. \end{cases}$$

Notemos que así $\Delta_k \equiv 0, 1 \pmod{4}$. La interpretación de esta definición es que el anillo de enteros de k es de la forma $\mathbb{Z}[\omega]$, donde el polinomio mínimo de ω es

$$f(x) = x^2 - d, \qquad \text{o bien} \qquad f(x) = x^2 - x + \frac{1-d}{4},$$

y en ambos casos el discriminante de este polinomio es Δ_k.

[1] Un detalle sutil es que no hemos probado que no pueda ocurrir que $k = \mathbb{Q}(\sqrt{d}) = \mathbb{Q}(\sqrt{d'})$ para distintos enteros d y d' libres de cuadrados. Si esto sucediera, entonces k tendría asociados dos discriminantes distintos. En realidad esto no puede ocurrir, y por eso es correcto asociar el discriminante a k en lugar de a d. Lo probamos en el ejemplo tras la definición 12.6, pero, para el uso que vamos a hacer de los discriminantes, el lector puede considerar que hemos definido un discriminante Δ_d en lugar de Δ_k. Lo mismo sucede con el carácter χ_k que definiremos seguidamente.

Si p es primo, la reducción de $f(x)$ en $\mathbb{Z}_p[x]$ factoriza como $\bar{f}(x) = (x-\bar{c})^2$ si y sólo si [2] $p \mid \Delta_k$.

En el caso de un primo impar $p \nmid \Delta_k$, tendremos que la reducción de $f(x)$ factoriza como $\bar{f}(x) = (x-\bar{c}_1)(x-\bar{c}_2)$ si y sólo si Δ_k es un resto cuadrático módulo p o, equivalentemente, $(\Delta_k/p) = 1$.

Para $p = 2$, si $2 \nmid \Delta_k$, tenemos que $f(x) = x^2 - x + (1-d)/4$, luego su reducción módulo 2 puede ser $x^2 + x = x(x+\bar{1})$, o bien $x^2 + x + \bar{1}$, que no tiene raíces en $\mathbb{Z}_2$, luego $\bar{f}(x)$ se descompone en dos factores distintos módulo 2 si y sólo si $d \equiv 1 \pmod 8$.

Como la condición $2 \nmid \Delta_k$ requiere que $d \equiv 1 \pmod 4$, luego $d \equiv 1, 5 \pmod 8$, la condición $d \equiv 1 \pmod 8$ es en realidad equivalente a que $\Delta_k \equiv \pm 1 \pmod 8$ (pues el signo negativo no puede darse), que a su vez puede expresarse en términos del símbolo de Jacobi como que $(2/\Delta_k) = 1$.

Vamos a ver que el símbolo de Jacobi nos permite reunir los dos casos en una condición común.

Definición 9.14 Si $k = \mathbb{Q}(\sqrt{d}\,)$ es un cuerpo cuadrático, definimos el *carácter* de k como la aplicación

$$\chi_k : U_{|\Delta_k|} \longrightarrow \{-1, 1\}$$

dada por $\chi_k(\bar{n}) = (\Delta_k/n) = (d/n)$, donde el representante n de la clase de restos se elige impar y positivo.

Siempre podemos elegir un representante impar positivo porque, si n es par y $\bar{n}$ está en $U_{|\Delta_k|}$ es que Δ_k es impar, por lo que sumando a n un múltiplo impar de Δ_k podemos pasar a un representante de su misma clase que cumpla lo requerido. La sexta propiedad del símbolo de Jacobi que probamos tras la definición 7.6 nos asegura que la definición de $\chi_k(\bar{n})$ no depende de la elección de dicho representante.

Notemos que el carácter χ_k está definido para todo cuerpo cuadrático, aunque su anillo de enteros no tenga factorización única, y las propiedades del símbolo de Jacobi implican trivialmente que $\chi_k(ab) = \chi_k(a)\chi_k(b)$.

Un poco más en general, podemos definir $\chi_k : \mathbb{Z} \longrightarrow \{-1, 0, 1\}$ mediante

$$\chi_k(n) = \begin{cases} \chi_k(\bar{n}) & \text{si } (n, \Delta_k) = 1, \\ 0 & \text{en caso contrario.} \end{cases}$$

Si Δ_k es impar, conviene observar que

$$\chi_k(2) = \left(\frac{2}{|\Delta_k|}\right).$$

[2] En principio, aquí hay que suponer que p es impar, porque la fórmula de la ecuación de segundo grado no vale en cuerpos de característica 2. No obstante, lo cierto es que también es cierto para $p = 2$, pues si $2 \mid \Delta_k$, entonces el polinomio es $x^2 - d$, que en $\mathbb{Z}_2$ es x^2 o bien $x^2 - 1 = (x-1)^2$, mientras que si $p \nmid \Delta_k$, el polinomio es $x^2 - x = x(x-1)$ o bien $x^2 + x + 1$, que no tiene raíces.

En efecto, tenemos que $\Delta_k \equiv 1 \pmod 4$. Tomemos un entero impar m tal que $2 + m\Delta_k > 0$. Si $\Delta_k < 0$, entonces $|\Delta_k| \equiv -1 \pmod 4$ y

$$\chi_k(2) = \left(\frac{\Delta_k}{2+m\Delta_k}\right) = \left(\frac{-1}{2+m\Delta_k}\right)\left(\frac{|\Delta_k|}{2+m\Delta_k}\right) = \left(\frac{2+m\Delta_k}{|\Delta_k|}\right) = \left(\frac{2}{|\Delta_k|}\right),$$

donde hemos usado que, por la ley de reciprocidad cuadrática, para invertir el segundo símbolo de Jacobi, siendo el "numerador" congruente con -1 módulo 4, hay que multiplicar precisamente por $(-1/2 + m\Delta_k)$, por lo que desaparece el primer símbolo de Legendre. Si $\Delta_k > 0$ el razonamiento es más sencillo.

Teorema 9.15 *Sea k un cuerpo cuadrático, sea $\mathbb{Z}[\omega]$ su anillo de enteros, sea $f(x)$ el polinomio mínimo de ω y sea $p > 0$ un primo racional. Sea $\bar{f}(x)$ la reducción de $f(x)$ en $\mathbb{Z}_p[x]$. Entonces:*

1. *Si $\chi_k(p) = 0$, entonces $\bar{f}(x) = (x - \bar{c})^2$, para cierto entero c.*
2. *Si $\chi_k(p) = 1$, entonces $\bar{f}(x) = (x - \bar{c}_1)(x - \bar{c}_2)$, donde $c_1 \not\equiv c_2 \pmod p$.*
3. *Si $\chi_k(p) = -1$, entonces $\bar{f}(x)$ es irreducible en $\mathbb{Z}_p[x]$.*

En particular, si $\mathbb{Z}[\omega]$ tiene factorización única, se cumple que p se ramifica, se escinde o se conserva primo en $\mathbb{Z}[\omega]$ si y sólo si $\chi_k(p) = 0, 1, -1$, respectivamente.

DEMOSTRACIÓN: La condición $\chi_k(p) = 0$ equivale a que $p \mid \Delta_k$, y ya hemos visto que esto equivale a que $\bar{f}(x)$ es un cuadrado en $\mathbb{Z}_p[x]$.

Igualmente, si p es impar, hemos probado que $\bar{f}(x)$ se descompone en dos factores primos distintos si y sólo si $p \nmid \Delta_k$ y $(\Delta_k/p) = 1$, lo cual equivale a que $\chi_k(p) = 1$.

Para $p = 2$ y $2 \nmid \Delta_k$, tenemos que

$$\chi_k(2) = \left(\frac{2}{|\Delta_k|}\right),$$

y también hemos probado que este símbolo de Jacobi vale 1 si y sólo si $\bar{f}(x)$ se descompone en dos factores primos distintos.

Como los tres casos son mutuamente excluyentes y hemos probado las dos primeras equivalencias, también tenemos la tercera. La última parte del teorema es consecuencia inmediata de 9.13. ∎

Si llamamos *clases de escisión* en $U_{|\Delta_k|}$ respecto de k a las clases de restos que cumplen $\chi_k(a) = 1$, la definición de χ_k implica inmediatamente las propiedades siguientes (que no requieren que el anillo de enteros de k tenga factorización única):

1. *Todo cuadrado en $U_{|\Delta_k|}$ es una clase de escisión.*
2. *El producto de clases de restos en $U_{|\Delta_k|}$ satisface:*

$$E \cdot E = E, \quad E \cdot N = N \cdot E = N, \quad N \cdot N = E,$$

donde E significa "clase de escisión" y N "clase que no es de escisión".

3. *La inversa de una clase de escisión es una clase de escisión.*

   Hay otras dos propiedades que no son inmediatas, pero que tampoco son difíciles de probar:

4. $\chi_k(-\bar{1}) = \begin{cases} 1 & \text{si } d > 0, \\ -1 & \text{si } d < 0. \end{cases}$

   En efecto, pongamos que $d = \epsilon 2^j m$, donde $\epsilon = \pm 1$ y $m > 0$ es impar. Entonces

   $$\chi_k(-\bar{1}) = \left(\frac{d}{4|d|-1}\right) = \left(\frac{\epsilon}{4|d|-1}\right)\left(\frac{2}{4|d|-1}\right)^j \left(\frac{m}{4|d|-1}\right).$$

   Si $j \neq 0$, entonces $j = 1$ (pues d es libre de cuadrados), luego $8 \mid 4|d|$, luego en cualquier caso

   $$\left(\frac{2}{4|d|-1}\right)^j = 1.$$

   Por otra parte,

   $$\left(\frac{m}{4|d|-1}\right) = (-1)^{(m-1)/2}\left(\frac{4|d|-1}{m}\right) = (-1)^{(m-1)/2}\left(\frac{-1}{m}\right) = 1.$$

   Por consiguiente

   $$\chi_k(-\bar{1}) = \left(\frac{\epsilon}{4|d|-1}\right) = \epsilon.$$

5. *El carácter χ_k toma los dos valores ± 1, por lo que hay exactamente $\phi(|\Delta_k|)/2$ clases de escisión.*

   En efecto, por la propiedad precedente podemos suponer que $d > 0$. Si $d = 2$ es fácil ver que $\chi_k(\bar{3}) = -1$. Si $d \neq 2$ podemos tomar un primo impar p tal que $d = pm$ y a su vez un entero u tal que $(u/p) = -1$. Por el teorema chino del resto existe un entero $n > 0$ tal que $n \equiv u$ (mód p), $n \equiv 1$ (mód $4m$) (pues, al ser d libre de cuadrados, $(p, 4m) = 1$). Notemos que $(n, d) = 1$.

   Entonces

   $$\chi_k(\bar{n}) = \left(\frac{d}{n}\right) = \left(\frac{p}{n}\right)\left(\frac{m}{n}\right) = \left(\frac{n}{p}\right)\left(\frac{m}{n}\right) = \left(\frac{u}{p}\right)\left(\frac{m}{1}\right) = -1.$$

   Fijado un n tal que $\chi_k(\bar{n}) = -1$, tenemos que $a \mapsto a \cdot \bar{n}$ hace corresponder las clases de escisión con las que no son de escisión, luego hay el mismo número de cada tipo. ∎

Ejemplo Para determinar el carácter de $\mathbb{Q}(\sqrt{6}\,)$ basta calcular

$$\chi_k(\bar{5}) = \left(\frac{6}{5}\right) = 1,$$

pues con esto ya sabemos que $\pm\bar{1}$ y $\pm\bar{5}$ son clases de escisión y, como ya son la mitad de las clases de U_{24}, no hay más. ∎

Es interesante observar que un fragmento de la parte final del teorema 9.15 no requiere la hipótesis de la factorización única:

Teorema 9.16 *Si K es un cuerpo cuadrático y $p \geq 2$ es un primo racional tal que $\chi_k(p) = -1$, entonces p es primo en el anillo de enteros de K y el anillo de clases de restos módulo p es un cuerpo con p^2 elementos.*

DEMOSTRACIÓN: Sea $\mathbb{Z}[\omega]$ el anillo de enteros de K y sea $q(x)$ el polinomio mínimo de ω. Sea $\bar{q}(x)$ el polinomio en $\mathbb{Z}_p[x]$ cuyos coeficientes son las clases módulo p de los coeficientes de $q(x)$. La hipótesis $\chi_k(p) = -1$ equivale a que $\bar{q}(x)$ no tiene raíces en $\mathbb{Z}_p$, luego el teorema 6.1 nos da un cuerpo k en el que $\bar{q}(x)$ tiene una raíz ω', cuyos elementos se expresan de forma única como $a+b\omega'$, con $a, b \in \mathbb{Z}_p$. En particular k tiene p^2 elementos.

Consideramos ahora la aplicación $f : \mathbb{Z}[\omega] \longrightarrow k$ dada por $a+b\omega \mapsto \bar{a}+\bar{b}\,\omega'$. Una comprobación rutinaria muestra que

$$f(\alpha+\beta) = f(\alpha) + f(\beta), \qquad f(\alpha\beta) = f(\alpha)f(\beta).$$

En efecto, para la suma es inmediato y, si $f(x) = x^2 + ux + v$, basta tener en cuenta que $\omega^2 = -u\omega - v$ y $\omega'^2 = -\bar{u}\omega' - \bar{v}$, por lo que la regla para calcular un producto arbitrario es la misma en ambos anillos.

Si $p \mid a + b\omega$, entonces $p \mid a$ y $p \mid b$, luego $f(a + b\omega) = 0$. Más en general, esto implica que si $\alpha \equiv \beta \pmod{p}$, entonces $p \mid \alpha - \beta$, luego $f(\alpha - \beta) = 0$, luego $f(\alpha) = f(\beta)$.

Por lo tanto, podemos definir $\bar{f} : \mathbb{Z}[\omega]_p \longrightarrow k$ mediante $\bar{\alpha} \mapsto f(\alpha)$, y obviamente se cumple que

$$\bar{f}(\bar{\alpha}+\bar{\beta}) = \bar{f}(\bar{\alpha}) + \bar{f}(\bar{\beta}), \quad \bar{f}(\bar{\alpha}\bar{\beta}) = \bar{f}(\bar{\alpha})\bar{f}(\bar{\beta}).$$

Más aún, como $\bar{f}(\overline{a+b\omega}) = \bar{a}+\bar{b}\omega'$, es obvio que todo elemento de k es imagen de algún elemento del anillo de clases de restos $\mathbb{Z}[\omega]_p$, pero éste tiene a lo sumo p^2 elementos (pues todo elemento de $\mathbb{Z}[\omega]_p$ puede expresarse como $\bar{a}+\bar{b}\bar{\omega}$, con $0 \leq a, b < p$) y k tiene exactamente p^2 elementos, luego necesariamente los elementos de $\mathbb{Z}[\omega]_p$ se corresponden biunívocamente con los de k, luego podemos identificar ambos anillos y, como k es un cuerpo, $\mathbb{Z}[\omega]_p$ también lo es.

El hecho de que $\mathbb{Z}[\omega]_p$ sea un cuerpo implica inmediatamente que p es primo, pues si $p \mid \alpha\beta$, entonces $\bar{\alpha} \cdot \bar{\beta} = 0$, luego $\bar{\alpha} = 0$ o bien $\bar{\beta} = 0$, luego $p \mid \alpha$ o $p \mid \beta$. ∎

Ejercicio: Sea $k = \mathbb{Q}(\sqrt{26})$, cuyo anillo de enteros no tiene factorización única. Comprobar que $\chi_k(3) = -1$ así como que el anillo de clases de restos módulo 3 de $\mathbb{Z}[\sqrt{26}]$ es un cuerpo de 9 elementos (calculando explícitamente la tabla del producto y comprobando que todo elemento no nulo tiene inverso).

Ejercicio: Sea $k = \mathbb{Q}(\sqrt{10})$, cuyo anillo de enteros no tiene factorización única. Comprobar que el anillo de clases de restos módulo 3 de $\mathbb{Z}[\sqrt{10}]$ tiene 9 elementos, pero no es un cuerpo.

9.6 Ejemplos y aplicaciones

La aplicación más elemental de los resultados que hemos obtenido es observar que ahora podemos extender los argumentos dados en los capítulos anteriores sobre los números de la forma x^2+y^2, x^2+2y^2 o x^2+3y^2 a muchos otros casos. Recordemos que el caso x^2-y^2 es mucho más elemental y lo analizamos al principio del capítulo IV.

Números de la forma x^2-2y^2 En principio, determinar si un número n es de la forma $n=x^2-2y^2$ es más complicado que determinar si es de la forma $n=x^2+2y^2$, pues en este segundo caso basta considerar todos los números naturales $0\le y\le\sqrt{n/2}$ y ver, para cada uno de ellos, si existe un x que cumpla la ecuación. En cambio, en el caso $n=x^2-2y^2$ no sabemos lo grandes que pueden ser unos valores de x e y que cumplan la ecuación.

Sin embargo, si reformulamos el problema como el de determinar si existen o no enteros de $\mathbb{Z}\big[\sqrt{2}\,\big]$ de norma n, la respuesta es muy sencilla. Repetimos el argumento: sea α un entero cuadrático y descompongámoslo en factores primos:

$$\alpha=\epsilon\pi_1^{e_1}\cdots\pi_r^{e_r},$$

donde los π_i son primos cuadráticos no asociados dos a dos y ϵ es una unidad. Entonces

$$\mathrm{N}(\alpha)=\pm\,\mathrm{N}(\pi_1)^{e_1}\cdots\mathrm{N}(\pi_r)^{e_r}.$$

Si llamamos $\pi_1,\dots,\pi_k$ a los divisores de primos racionales que se conservan, de modo que $|\,\mathrm{N}(\pi_i)|=p_i^2$, y $\pi_{k+1},\dots,\pi_r$ a los divisores de primos racionales p_i que se ramifican o se escinden, de modo que $|\,\mathrm{N}(\pi_i)|=p_i$, tenemos que

$$|\,\mathrm{N}(\alpha)|=p_1^{2e_1}\cdots p_k^{2e_k}p_{k+1}^{e_{k+1}}\cdots p_r^{e_r}.$$

Vemos que un número natural n es de esta forma si y sólo si, en su descomposición en factores primos, los primos que se conservan en $\mathbb{Z}\big[\sqrt{2}\,\big]$ aparecen con exponente par o, equivalentemente, si los primos que dividen a n con exponente impar se ramifican o se escinden en $\mathbb{Z}\big[\sqrt{2}\,\big]$.

Si $k=\mathbb{Q}\big(\sqrt{2}\,\big)$ tenemos que $\Delta_k=8$, luego χ_k está definido sobre el grupo de unidades $U_8=\{\pm\bar1,\pm\bar3\}$. Sabemos a priori que $\chi_k(\pm\bar1)=1$, luego las clases $\pm\bar1$ son todas las clases de escisión y 2 es el único primo que se ramifica.

Pero queda pendiente el problema del signo de la norma. En principio, si un número entero n cumple que los primos que lo dividen con exponente impar cumplen $p=2$ o $p\equiv\pm1\pmod 8$, con esto sabemos que $\pm n=x^2-2y^2$, pero no sabemos si concretamente n es de esta forma, o sólo lo es $-n$.

Ahora bien, esto no es problema, pues en $\mathbb{Z}\big[\sqrt{2}\,\big]$ existen unidades de norma negativa, por ejemplo $\eta=1+\sqrt{2}$. Por lo tanto, si

$$n=x^2-2y^2=\mathrm{N}(x+y\sqrt{2}),$$

entonces

$$-n=\mathrm{N}(\eta)\,\mathrm{N}(x+y\sqrt{2})=\mathrm{N}((1+\sqrt{2})(x+y\sqrt{2}))=$$
$$\mathrm{N}((x+2y)+(x+y)\sqrt{2})=(x+2y)^2-2(x+y)^2.$$

Así pues, n es de la forma $x^2 - 2y^2$ si y sólo si lo es $-n$ y podemos concluir:

Un número entero es de la forma $x^2 - 2y^2$ si y sólo si los primos que lo dividen con exponente impar cumplen $p = 2$ o $p \equiv \pm 1$ (mód 8).

■

Números de la forma $x^2 - 3y^2$ La situación es ligeramente distinta para el caso de los números de la forma $x^2 - 3y^2$. Consideramos $k = \mathbb{Q}(\sqrt{3})$, cuyo anillo de enteros es $\mathbb{Z}[\sqrt{3}]$, y así lo que queremos es caracterizar las normas de los enteros de k. Tenemos que $\Delta_k = 12$, luego χ_k está definido sobre el grupo de unidades $U_{12} = \{\pm\bar{1}, \pm\bar{5}\}$ y es inmediato que las clases de escisión son ± 1.

Exactamente el mismo argumento del apartado precedente nos permite concluir que un entero n cumple que $\pm n = x^2 - 3y^2$ si y sólo si los primos que lo dividen con exponente impar cumplen $p = 2, 3$ o bien $p \equiv \pm 1$ (mód 12). Sin embargo, ahora la cuestión del signo no puede ser soslayada.

La razón es que la unidad fundamental de k es $\eta = 2 + \sqrt{3}$ y $\mathrm{N}(\eta) = 1$, por lo que todas las unidades tienen norma 1, y no es posible convertir una expresión de n en la forma $x^2 - 3y^2$ en otra para $-n$.

Por ejemplo, $3 = \sqrt{3}^2$ se ramifica, luego todos los primos de norma ± 3 son asociados a $\pi = \sqrt{3}$, que tiene norma -3, luego todos los primos de norma ± 3 tienen, de hecho, norma -3.

Similarmente, $2 = -(1 + \sqrt{3})(1 - \sqrt{3})$, pero los dos factores son asociados (tienen que serlo, porque sabemos que 2 se ramifica). Dividimos

$$\frac{1 - \sqrt{3}}{1 + \sqrt{3}} = \frac{(1 - \sqrt{3})^2}{-2} = -2 + \sqrt{3},$$

de modo que

$$2 = (2 - \sqrt{3})(1 + \sqrt{3})^2.$$

Todos los primos de norma ± 2 son asociados a $1 + \sqrt{3}$, luego todos ellos tienen norma -2.

En general, si vamos calculando números de la forma $x^2 - 3y^2$ y nos quedamos con los primos que aparecen, obtenemos:

$$-2, \quad -3, \quad -11, \quad 13, \quad -23, \quad 37, \quad -47,$$

$$-59, \quad 61, \quad -71, \quad 73, \quad -83, \quad 97, \quad \ldots$$

No es difícil conjeturar cuál es la situación:

Un primo $p > 0$ es de la forma $p = x^2 - 3y^2$ si y sólo si cumple $p \equiv 1$ (mód 12), mientras que $-p = x^2 - 3y^2$ si y sólo si $p = 2, 3$ o bien $p \equiv -1$ (mód 12).

Podemos probarlo mediante "fuerza bruta" considerando todas las posibilidades módulo 12:

x, y (mód 12)	0	±1	±2	±3	±4	±5	6
x^2 (mód 12)	0	1	4	-3	4	1	0
$-3y^2$ (mód 12)	0	-3	0	-3	0	-3	0

x^2-3y^2	0	1	4	-3
0	0	1	4	-3
-3	-3	-2	1	0

Vemos así que si un primo $p > 3$ cumple $p = x^2 - 3y^2$, por una parte tiene que cumplir $p \equiv \pm1$ (mód 1)2, pero la tabla precedente muestra que el signo negativo es imposible, luego así tenemos probado:

> *Un primo $p > 0$ cumple $p = x^2 - 3y^2$ si y sólo si $p \equiv 1$ (mód 12), mientras que $-p = x^2 - 3y^2$ si y sólo si cumple $p = 2, 3$ o bien $p \equiv -1$ (mód 12).*

Con esto es fácil razonar la situación general. Dejamos al lector que justifique la afirmación siguiente:

> *Un número natural $n \geq 1$ es de la forma $n = x^2 - 3y^2$ si y sólo si los primos que lo dividen con exponente impar cumplen $p = 2, 3$ o $p \equiv \pm1$ (mód 12) y el número de los primos que cumplen $p = 2, 3$ o $p \equiv -1$ (mód 12) es par.* ∎

Ejercicio: Caracterizar los enteros de la forma $n = x^2 - 4y^2$.

Números de la forma $x^2 - 5y^2$ Los enteros de esta forma son las normas de los elementos del anillo $\mathbb{Z}[\sqrt{5}]$, pero hay que tener en cuenta que éste no es el anillo de enteros del cuerpo $k = \mathbb{Q}(\sqrt{5})$, sino que éste es $\mathbb{Z}[\omega]$, donde $\omega = (1 + \sqrt{5})/2$. Además,

$$\mathrm{N}(x + y\omega) = x^2 + xy - y^2,$$

por lo que, en principio, con las técnicas que estamos empleando, podemos aspirar a identificar los enteros de la forma $n = x^2 + xy - y^2$, no los de la forma $n = x^2 - 5y^2$.

Ahora bien, observamos que todo elemento de $\mathbb{Z}[\omega]$ tiene un asociado en $\mathbb{Z}[\sqrt{5}]$ con la misma norma. En efecto, basta tener en cuenta que

$$\omega(x + y\omega) = \frac{x + 3y}{2} + \frac{x + y}{2}\sqrt{5}, \qquad \omega^2(x + y\omega) = \frac{3x + 4y}{2} + \frac{x + 2y}{2}\sqrt{5},$$

de donde $\omega^i(x + y\omega) \in \mathbb{Z}[\sqrt{5}]$ para algún $i = 0, 1, 2$. Además, $\omega^3 = 2 + \sqrt{5}$ cumple $\mathrm{N}(\omega^3) = -1$, luego multiplicando por ω^3 si es necesario podemos asegurar que un asociado de $x + y\omega$ está en $\mathbb{Z}[\sqrt{5}]$ y tiene la misma norma.

En otras palabras, tenemos que las normas de los elementos de $\mathbb{Z}[\omega]$ son las mismas que las normas de los elementos de $\mathbb{Z}\big[\sqrt{5}\,\big]$, así como que hay elementos de norma n si y sólo si los hay de norma $-n$, por lo que el signo de la norma no debe preocuparnos. A partir de aquí, todo es pura rutina. Tenemos que $\Delta_k = 5$, y las clases de escisión en U_5 son $\pm\bar{1}$, luego:

> *Un entero n es de la forma $n = x^2 - 5y^2$ (y también de la forma $n = x^2+xy-y^2$) si y sólo si los primos que lo dividen con exponente impar cumplen $p = 2, 5$ o $p \equiv \pm 1$ (mód 5).*

∎

Números de la forma $x^2 + 5y^2$ Si intentamos caracterizar los enteros de la forma $n = x^2+5y^2$ nos encontramos con una dificultad crucial, y es que el anillo de enteros del cuerpo $k = \mathbb{Q}\big(\sqrt{-5}\,\big)$, es decir, $\mathbb{Z}\big[\sqrt{-5}\,\big]$, no tiene factorización única.

Vamos a "olvidarnos" de este hecho y veamos cómo serían los números de la forma $x^2 + 5y^2$ si $\mathbb{Z}\big[\sqrt{-5}\,\big]$ fuera un dominio de factorización única. Como $\Delta_k = -20$, tenemos que calcular las clases de escisión del grupo

$$U_{20} = \{\pm\bar{1}, \pm\bar{3}, \pm\bar{7}, \pm\bar{9}\}.$$

Una de ellas es $\bar{1}$ y otra $\bar{9}$, porque es un cuadrado, si calculamos, por ejemplo,

$$\chi_k(\bar{3}) = \left(\frac{-5}{3}\right) = \left(\frac{1}{3}\right) = 1,$$

concluimos que $\bar{3}$ también es una clase de escisión, así como $\bar{9}\cdot\bar{3} = \bar{7}$. Por lo tanto, las clases de escisión son $\bar{1}, \bar{3}, \bar{7}, \bar{9}$. Por lo tanto, un número natural n "debería" ser de la forma $x^2 + 5y^2$ si y sólo si los primos que lo dividen con exponente impar cumplen

$$p = 2, 5 \qquad \text{o} \qquad p \equiv 1, 3, 7, 9 \text{ (mód 20)}.$$

Ahora contrastemos esto con "la realidad". La tabla siguiente muestra los primeros números de la forma $x^2 + 5y^2$:

1	4	**5**	6	9	14	16	20	21	24	25	**29**	30	36	**41**	45
46	49	54	56	**61**	64	69	70	80	81	84	86	**89**	94	96	100
101	105	**109**	116	120	121	125	126	129	134	141	144	145	**149**	150	161
164	166	169	174	180	**181**	184	189	196	201	205	206	214	216	224	225
229	230	**241**	244	245	246	249	254	256	261	**269**	270	276	280	**281**	289
294	301	305	309	320	321	324	326	329	334	336	344	345	**349**	350	356
361	366	369	376	381	384	**389**	400	**401**	404	405	406	**409**	414	420	**421**
430	436	441	445	446	**449**	454	**461**	464	469	470	480	484	486	489	500

Los primos que aparecen en la lista son el 5 (pero no el 2) y los que cumplen $p \equiv 1, 9$ (mód 20).

Ejercicio: Probar que un primo que cumpla $p \equiv 3, 7$ (mód 20) no puede ser de la forma $p = x^2 + 5y^2$.

Sin embargo, vemos que "los primos que faltan" sí que aparecen como divisores de otros números de la tabla. En realidad, para evitar una obviedad tenemos que fijarnos en la parte libre de cuadrados de cada número (pues cualquier primo aparece como factor de un número de la forma $(kx)^2+5(ky)^2$). La tabla siguiente muestra la parte libre de cuadrados de cada número de la tabla anterior (en negrita están los números que ya eran libres de cuadrados):

1	1	**5**	**6**	1	**14**	1	5	**21**	6	1	**29**	**30**	1	**41**	5
46	1	6	14	**61**	1	**69**	**70**	5	1	21	**86**	**89**	**94**	6	1
101	**105**	**109**	29	30	1	5	14	**129**	**134**	**141**	1	**145**	**149**	6	**161**
41	**166**	1	**174**	5	**181**	46	21	1	**201**	**205**	**206**	**214**	6	14	1
229	**230**	**241**	61	5	**246**	**249**	**254**	1	**261**	**269**	**270**	**276**	**280**	**281**	1
294	**301**	**305**	**309**	5	**321**	1	**326**	**329**	**334**	21	86	**345**	**349**	14	89
1	**366**	41	94	**381**	6	**389**	1	**401**	101	5	**406**	**409**	**414**	**420**	**421**
430	109	1	**445**	**446**	**449**	**454**	**461**	29	**469**	**470**	30	1	**486**	**489**	5

Esta tabla invita a conjeturar que un número es de la forma x^2+5y^2 si y sólo si lo es su parte libre de cuadrados (una implicación es obvia, pero la otra no). Si ahora nos fijamos en los factores primos de estos números, resultan ser:

$$2, 3, 5, 7, 23, 29, 41, 43, 47, 61, 67, 83, 89, 101, 103, 107, 109, 127, 149, 163, 167, \ldots$$

Y ésta —aparte de 2 y 5— no es sino la lista de todos los primos que cumplen $p \equiv 1, 3, 7, 9 \pmod{20}$.

Así pues, vemos que los números de la forma x^2+5y^2 tienen su parte libre de cuadrados formada por "los primos que deberían ser", pero no todas las combinaciones de dichos primos son válidas. Tal vez el lector pueda conjeturar algo sobre qué combinaciones de primos (libres de cuadrados) dan lugar a números de la forma x^2+5y^2. Por ejemplo, Euler observó (conjeturó) lo siguiente, aunque no pudo probarlo:

> *Si p y q son primos tales que $p, q \equiv 3, 7 \pmod{20}$, su producto pq es de la forma x^2+5y^2.*

De momento no podemos demostrar ninguna de estas conjeturas, pero resulta evidente que todas ellas están relacionadas de algún modo con la aritmética del anillo $\mathbb{Z}\big[\sqrt{-5}\,\big]$. ∎

Una conjetura de Ramanujan Como aplicación de la factorización única en $\mathbb{Q}\big(\sqrt{-7}\,\big)$ mostramos una prueba debida a Nagell de una conjetura de Ramanujan:

> *Las únicas soluciones enteras de la ecuación $x^2+7=2^n$ son:*
>
> $$(x,n) \quad = \quad (\pm 1, 3), \quad (\pm 3, 4), \quad (\pm 5, 5), \quad (\pm 11, 7), \quad (\pm 181, 15).$$

Demostración: Claramente, x tiene que ser impar, y podemos suponer que es positivo. Supongamos primero que $n = 2m$. Entonces podemos factorizar

$$(2^m + x)(2^m - x) = 7,$$

de donde $2^m + x = 7$ y $2^m - x = 1$. Sumando, $2^{m+1} = 8$, luego $m = 2$, $n = 4$, $x = 3$.

Sea ahora n impar. El caso $n = 3$ lleva a la solución $(1, 3)$. Supongamos que $n > 3$.

Trabajamos en el anillo de enteros de $\mathbb{Q}\big(\sqrt{-7}\,\big)$. La descomposición en factores primos de 2 es la siguiente:

$$2 = \left(\frac{1+\sqrt{-7}}{2}\right)\left(\frac{1-\sqrt{-7}}{2}\right).$$

Como x es impar, es fácil ver que x^2+7 es múltiplo de 4. Tomando $m = n-2$ podemos reescribir la ecuación del siguiente modo:

$$\frac{x^2+7}{4} = 2^m$$

y podemos factorizarla así en $\mathbb{Q}\big(\sqrt{-7}\,\big)$:

$$\left(\frac{x+\sqrt{-7}}{2}\right)\left(\frac{x-\sqrt{-7}}{2}\right) = \left(\frac{1+\sqrt{-7}}{2}\right)^m \left(\frac{1-\sqrt{-7}}{2}\right)^m.$$

Ninguno de los primos de la derecha divide a la vez a los dos factores de la izquierda, pues entonces dividiría a su diferencia, $\sqrt{-7}$, pero las normas correspondientes no se dividen.

Por la unicidad de la factorización (y teniendo en cuenta que las unidades son ± 1), podemos concluir que

$$\frac{x \pm \sqrt{-7}}{2} = \pm\left(\frac{1 \pm \sqrt{-7}}{2}\right)^m,$$

donde esta expresión representa a dos ecuaciones de las que desconocemos los signos adecuados, pero, en cualquier caso, la ecuación para $\frac{x+\sqrt{-7}}{2}$ debe tener a la derecha un primo y la de $\frac{x-\sqrt{-7}}{2}$ debe tener el otro, mientras que la unidad ± 1 debe ser la misma en ambas ecuaciones. Al restarlas queda

$$\pm\sqrt{-7} = \left(\frac{1+\sqrt{-7}}{2}\right)^m - \left(\frac{1-\sqrt{-7}}{2}\right)^m.$$

Ahora probamos que el signo ha de ser negativo. Supongamos que es positivo. Llamemos $a = \frac{1+\sqrt{-7}}{2}$ y $b = \frac{1-\sqrt{-7}}{2}$. Estamos suponiendo que

$$a^m - b^m = a - b.$$

Como $a + b = 1$ y $ab = 2$, tenemos lo siguiente:

$$a^2 = (1-b)^2 = 1 - 2b + b^2 = 1 - ab^2 + b^2 \equiv 1 \pmod{b^2}$$

Por lo tanto $a^m = a(a^2)^{(m-1)/2} \equiv a \pmod{b^2}$, y así

$$a - b = a^m - b^m \equiv a - 0 \pmod{b^2},$$

es decir, $b^2 \mid b$, lo cual es imposible.

Así pues, $-2^m\sqrt{-7} = \big(1+\sqrt{-7}\big)^m - \big(1-\sqrt{-7}\big)^m$. Al desarrollar por el teorema del binomio de Newton se cancelan los términos pares y queda

$$-2^m\sqrt{-7} = 2\binom{m}{1}\sqrt{-7} + 2\binom{m}{3}\big(\sqrt{-7}\big)^3 + \cdots + 2\binom{m}{m}\big(\sqrt{-7}\big)^m.$$

Sacamos factor común 2 y concluimos que $-2^{m-1} \equiv m \pmod 7$.

Como $2^6 \equiv 1 \pmod 7$, es claro que si un número m cumple esta congruencia, también lo cumplen todos los congruentes con m módulo 42.

Si examinamos todos los números entre 0 y 41, vemos que los únicos que cumplen la congruencia son $m = 3$, 5 y 13. El teorema estará probado si demostramos que dos soluciones de la ecuación original (en el caso que estamos estudiando) no pueden ser congruentes módulo 42, pues entonces las únicas soluciones posibles serán $n = 5$, 7, 15, con las que se corresponden $x = 5$, 11, 181.

Supongamos que m y m' son soluciones de la ecuación (en realidad queremos decir que $m+2$ y $m'+2$ lo son). Supongamos que $m < m'$ y $m \equiv m' \pmod{42}$. Sea 7^l la mayor potencia de 7 que divide a $m'-m$. Entonces

$$a^{m'} = a^m a^{m'-m} = a^m \left(\frac{1}{2}\right)^{m'-m} \big(1+\sqrt{-7}\big)^{m'-m}. \tag{9.1}$$

Como $\phi(7^{l+1}) = 6\cdot 7^l$, el teorema de Fermat implica que $2^{6\cdot 7^l} \equiv 1 \pmod{7^{l+1}}$, y como $6\cdot 7^l \mid m'-m$, tenemos

$$2^{m'-m} \equiv 1 \pmod{7^{l+1}}. \tag{9.2}$$

Una sencilla inducción (en la que se usa el teorema 3.15) prueba que

$$\big(1+\sqrt{-7}\big)^{7^l} = 1 + 7^l\sqrt{-7}\,(1+7\alpha),$$

para cierto entero α. Al elevar después a $(m'-m)/7^l$ y desarrollar, obtenemos que

$$(1+\sqrt{-7})^{m'-m} \equiv 1 + (m'-m)\sqrt{-7} \pmod{7^{l+1}}. \tag{9.3}$$

Ahora pasamos la potencia de 2 al primer miembro de (9.1), tomamos congruencias módulo 7^{l+1} en el anillo de enteros de $\mathbb{Q}\big(\sqrt{-7}\big)$ y sustituimos (9.2) y (9.3), de lo que resulta

$$a^{m'} \equiv a^m + a^m(m'-m)\sqrt{-7} \pmod{7^{l+1}}. \tag{9.4}$$

De nuevo por inducción se prueba que $a^m \equiv \frac{1+m\sqrt{-7}}{2^m} \pmod 7$, luego podemos escribir $a^m = \frac{1+m\sqrt{-7}}{2^m} + 7\alpha$ para un cierto entero α. Como $7^l \mid m'-m$, al sustituir en (9.4) queda

$$a^{m'} \equiv a^m + \frac{m'-m}{2^m}\sqrt{-7} \pmod{7^{l+1}}. \tag{9.5}$$

Conjugando:

$$b^{m'} \equiv b^m - \frac{m'-m}{2^m}\sqrt{-7} \text{ (mód } 7^{l+1}). \tag{9.6}$$

Pero al ser m y m' soluciones de la ecuación, tenemos

$$a^m - b^m = -\sqrt{-7} = a^{m'} - b^{m'},$$

luego al restar (9.5) menos (9.6) llegamos a que

$$(m'-m)\sqrt{-7} \equiv 0 \text{ (mód } 7^{l+1}).$$

Por lo tanto, la multiplicidad del primo $\sqrt{-7}$ en $m'-m$ es al menos $2l+1$, pero ésta ha de ser el doble de la multiplicidad de 7 (visto como primo de $\mathbb{Z}$) en $m'-m$, luego la multiplicidad de $\sqrt{-7}$ es al menos $2l+2$ y así $7^{l+1} \mid m'-m$, en contra de la elección de l. ∎

9.7 El test de Lucas-Lehmer

En la sección 6.4 señalamos que Édouard Lucas demostró que el número de Fermat F_6 es compuesto. Previamente había demostrado que lo mismo le ocurre al número de Mersenne M_{67}, en contra de lo que había afirmado Mersenne, pero su mayor proeza la obtuvo en 1876, cuando demostró que el número de Mersenne

$$M_{127} = 170\,141\,183\,460\,469\,231\,731\,687\,303\,715\,884\,105\,727$$

es primo. Hasta ese momento, el mayor primo conocido era M_{31}, establecido por Euler hacía más de un siglo, y M_{127} conservaría el puesto durante 75 años. Es el mayor número cuya primalidad se ha comprobado sin la ayuda de un ordenador. Para lograrlo utilizó una técnica totalmente diferente de las empleadas hasta entonces:[3]

Teorema 9.17 (Test de Lucas-Lehmer) *Consideremos la sucesión dada por*

$$s_0 = 4, \qquad s_{n+1} = s_n^2 - 2.$$

Si p es un primo impar, entonces $M_p = 2^p - 1$ es primo si y sólo si $M_p \mid s_{p-2}$.

Demostración: Consideremos $\eta = 2+\sqrt{3}$ y observemos que $s_n = \eta^{2^n} + \bar{\eta}^{2^n}$.

En efecto, $\eta + \bar{\eta} = 4 = s_0$. Si la relación es cierta para n, entonces

$$s_{n+1} = s_n^2 - 2 = (\eta^{2^n} + \bar{\eta}^{2^n})^2 - 2 =$$

$$\eta^{2^{n+1}} + \bar{\eta}^{2^{n+1}} + 2(\eta\bar{\eta})^{2^n} - 2 = \eta^{2^{n+1}} + \bar{\eta}^{2^{n+1}}.$$

[3]El test de Pépin que hemos visto en la sección 6.4 es similar, pero es un año posterior, pues Pépin lo publicó en 1877.

Notemos también que $\bar{\eta} = \eta^{-1}$. Veamos ahora que

$$M_p \mid s_{p-2} \quad \text{si y sólo si} \quad \eta^{2^{p-1}} \equiv -1 \text{ (mód } M_p).$$

En efecto, $M_p \mid s_{p-2}$ si y sólo si $\eta^{2^{p-2}} + \bar{\eta}^{2^{p-2}} \equiv 0$ (mód M_p), si y sólo si

$$\eta^{2^{p-2}} \equiv -\eta^{-2^{p-2}} \text{ (mód } M_p),$$

y esto equivale a que $\eta^{2^{p-1}} \equiv -1$ (mód M_p).

Ahora veamos que $\eta^{2^{p-1}} \equiv -1$ (mód M_p) equivale a que M_p sea primo.

Supongamos que se da la congruencia. Si M_p no es primo, entonces tiene un divisor primo $q < \sqrt{M_p}$. Como M_p es impar, tiene que ser $q \geq 3$. Sea π un divisor primo de q en $\mathbb{Z}\big[\sqrt{3}\,\big]$. Entonces

$$\eta^{2^{p-1}} \equiv -1 \text{ (mód } \pi),$$

luego $\eta^{2^p} \equiv 1$ (mód π), luego η tiene orden 2^p módulo π. Este orden tiene que ser un divisor de el grupo de unidades del cuerpo de restos módulo π, que tiene orden $q-1$ o bien q^2-1, en cualquiera de los dos casos $2^p \mid q^2-1 = (q+1)(q-1)$.

Pero uno de los dos números $q \pm 1$ tiene que cumplir $q \pm 1 \equiv 2$ (mód 4), de modo que 2 lo divide con exponente 1, luego podemos concluir que $2^{p-1} \mid q \pm 1$. En particular $2^{p-1} \leq q + 1$, luego

$$M_p = 2^p - 1 \leq 2q + 2 - 1 = 2q + 1 < 2q + q = 3q \leq q^2,$$

luego $\sqrt{M_p} < q$ y tenemos una contradicción.

Supongamos ahora que M_p es primo. Como $p \geq 3$, tenemos que

$$M_p = 2^p - 1 \equiv -1 \text{ (mód } 8).$$

Consecuentemente,

$$-1 = \left(\frac{-2}{M_p}\right) \equiv (-2)^{(M_p-1)/2} \text{ (mód } M_p),$$

donde hemos usado las leyes suplementarias de la ley de reciprocidad cuadrática y el criterio de Euler 5.8. Sabemos que 2 se escinde en $\mathbb{Z}\big[\sqrt{3}\,\big]$. Concretamente, si $\pi = 1 + \sqrt{3}$, tenemos que

$$\pi\bar{\pi} = -2,$$

pero los dos factores son asociados. Concretamente,

$$\frac{\bar{\pi}}{\pi} = \frac{1-\sqrt{3}}{1+\sqrt{3}} = \frac{(1-\sqrt{3})^2}{-2} = -2 + \sqrt{3} = -\bar{\eta},$$

luego

$$2 = \bar{\eta}\pi^2.$$

Usando de nuevo el criterio de Euler:

$$\pi^{M_p} = (1+\sqrt{3})^{M_p} \equiv 1+\sqrt{3}^{M_p} \equiv 1+3^{(M_p-1)/2}\sqrt{3} \equiv 1+\left(\frac{3}{M_p}\right)\sqrt{3} \text{ (mód } M_p).$$

Pero

$$\left(\frac{3}{Mp}\right) = -\left(\frac{M_p}{3}\right) = -1,$$

pues $M_p = 2^p - 1 \equiv (-1)^p - 1 \equiv -2 \equiv 1$ (mód 3). Por consiguiente, tenemos que

$$\pi^{M_p} \equiv 1-\sqrt{3} = \bar{\pi} = -\pi\bar{\eta} \text{ (mód } M_p),$$

luego $\pi^{M_p-1} \equiv -\bar{\eta}$ (mód M_p).

De aquí obtenemos que

$$-1 \equiv (-2)^{(M_p-1)/2} = (\pi\bar{\pi})^{(M_p-1)/2} \equiv (\pi\pi^{M_p})^{(M_p-1)/2}$$

$$= \pi^{(M_p+1)(M_p-1)/2} \equiv (\pi^{M_p-1})^{(M_p+1)/2} \equiv \bar{\eta}^{2^{p-1}} \text{ (mód } M_p),$$

luego también $\eta^{2^{p-1}} \equiv -1$ (mód M_p). ∎

Ejemplo En la sección 3.5 probamos que el número de Mersenne M_{13} es primo. Ahora vamos mostrar los cálculos que hay que hacer para probar que M_{17} también lo es usando el criterio de Lucas-Lehmer. Para entender el procedimiento que emplearemos en la práctica (que es esencialmente el mismo que usó Lucas para probar que M_{127} es primo), consideraremos primero el caso de $M_5 = 31$.

En general, el primer hecho esencial es que no necesitamos calcular los términos de la sucesión s_n, que no tardan en volverse astronómicos, sino que basta calcular sus restos módulo el M_p cuya primalidad queramos investigar. Análogamente al tratamiento de los primos de Fermat que vimos en la sección 6.4, Lucas trabajó en base 2. Podemos partir de $s_1 = 4^2 - 2 = 14 = 1110_2$. Para calcular su cuadrado con el algoritmo usual de la multiplicación, el proceso es el principio del cálculo de la izquierda:

$$\begin{array}{r} 1110 \\ \times\ 1110 \\ \hline 1110 \\ 1110 \\ +\ 1110 \\ \hline 1232100 \\ 110|00100 \\ +\ 110 \\ \hline 1010 \\ -\ 10 \\ \hline 1000 \end{array} \qquad \begin{array}{r} 01110 \\ \hline 11100 \\ 11001 \\ 10011 \\ \hline 3211\mathbf{0} \\ 10|00110 \\ 10 \\ \hline 1000 \end{array}$$

Observemos que hemos sumado las "unidades", "decenas", etc. binarias para obtener una primera representación impropia del resultado como

$$1232100_2 = 1\cdot 2^6 + 2\cdot 2^5 + 3\cdot 2^4 + 2\cdot 2^3 + 1\cdot 2^2,$$

que se transforma en 11000100_2 sin más que sustituir cada dos "unidades" de un orden determinado por una "unidad" del orden siguiente:

$$1232100_2 = 1240100_2 = 1400100_2 = 3000100_2 = 11000100_2$$

Lucas trabajaba poniendo cuentas en una cuadrícula de celdas, con lo que este proceso es totalmente elemental en dicho contexto. Pero ahora observamos que $2^p \equiv 1 \pmod{M_p}$. En nuestro caso, $2^5 \equiv 1 \pmod{M_5}$, por lo que un dígito 1 con valor 2^5 equivale a un dígito 1 con valor 2^0, uno en con valor 2^6 equivale a uno con valor $2^2 1$, etc. En la práctica, esto supone que el 110 que ha quedado a partir de la quinta posición puede trasladarse al principio, con lo que el número se reduce a 1010_2. Esto es $s_1^2 \pmod{M_5}$ y, si le restamos $2 = 10_2$, obtenemos $s_2 = 1000_2$.

Ahora bien, todo este proceso se simplifica drásticamente (al menos cuando lo realizamos con primos mayores) si tenemos en cuenta que los dígitos posteriores al quinto pueden reducirse en cualquier momento del cálculo, de modo que nunca necesitamos trabajar con más de cinco dígitos binarios.

A la derecha hemos realizado el mismo cálculo pero trasladando inmediatamente cada dígito que debería ocupar una posición posterior a la quinta. Además, en el momento de realizar la suma correspondiente a la multiplicación, en lugar de sumar dos unidades, hemos puesto 0 descontando así ya en ese momento las 2 unidades que hay que restar para calcular $s_1^2 - 2$. Notemos que, al reducir la expresión binaria impropia, hemos escrito dos cifras en posiciones posteriores a la quinta, pero ha sido sólo para que el cálculo quede más claro sobre el papel. Si operamos con cuentas sobre celdas, podemos situarlos inmediatamente en su posición reducida.

Ahora es inmediato que $s_2^2 = 1000000_2 = 10_2$, por lo que $s_3 = 10_2 - 10_2 = 0$ (módulo M_5), y esto prueba que $M_5 = 31$ es primo.

La tabla siguiente resume el resultado de aplicar este mismo proceso al caso de $M_{17} = 131\,071$:

n	s_n	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
0	4															1	0	0
1	14														1	1	1	0
2	194										1	1	0	0	0	0	1	0
3	37 634			1	0	0	1	0	1	1	0	0	0	0	0	0	1	0
4	95 799	1	0	1	1	1	0	1	1	0	0	0	1	1	0	1	1	1
5	119 121	1	1	1	0	1	0	0	0	1	0	1	0	1	0	0	0	1
6	66 179	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	1
7	53 645		1	1	0	1	0	0	0	1	1	0	0	0	1	1	0	1
8	122 218	1	1	1	0	1	1	1	0	1	0	1	1	0	1	0	1	0
9	126 220	1	1	1	1	0	1	1	0	1	0	0	0	0	1	1	0	0
10	70 490	1	0	0	0	1	0	0	1	1	0	1	0	1	1	0	1	0
11	69 559	1	0	0	0	0	1	1	1	1	1	0	1	1	0	1	1	1
12	99 585	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1
13	78 221	1	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	1
14	130 559	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
15	0																	0

Hemos indicado el valor en base 10 del resto de s_n módulo M_{17}, pero no es necesario para el cálculo. La tabla siguiente detalla los cálculos necesarios para pasar de s_{10} a s_{11}. En primer lugar elevamos s_{10} al cuadrado, lo cual en la práctica significa permutar cíclicamente sus cifras para que el 0 final (en negrita en la tabla) vaya quedando debajo de cada 1 de s_{10}. Luego sumamos los unos de cada columna para obtener una representación impropia de s_{10}^2 en base 2 y, de paso, restamos ya 2 unidades, por lo que en la columna de unidades hemos puesto 2 (en negrita) en vez del 4 que habría que poner para calcular el cuadrado. A continuación hemos detallado el proceso de reducción a una expresión binaria propia. Hay que razonar de derecha a izquierda:

		16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
	s_{10}	1	0	0	0	1	0	0	1	1	0	1	0	1	1	0	1	**0**
		0	0	0	1	0	0	1	1	0	1	0	1	1	0	1	**0**	1
		0	1	0	0	1	1	0	1	0	1	1	0	1	**0**	1	0	0
		1	0	0	1	1	0	1	0	1	1	0	1	**0**	1	0	0	0
		0	1	1	0	1	0	1	1	0	1	**0**	1	0	0	0	1	0
		1	0	1	0	1	1	0	1	**0**	1	0	0	0	1	0	0	1
		0	1	0	1	1	0	1	**0**	1	0	0	0	1	0	0	1	1
		1	1	0	1	**0**	1	0	0	0	1	0	0	1	1	0	1	0
		0	1	0	0	0	1	0	0	1	1	0	1	0	1	1	0	1
		3	5	2	4	5	4	4	4	3	7	1	4	4	4	3	3	**2**
1	1	4/5	6	3/4	5/6	6	5	5	5	3	8/9	5	5	5	5/6	3	4	2
1	1	1	0	0	0	0	1	1	1	1	1	0	1	1	0	1	0	0
	s_{11}	1	0	0	0	0	1	1	1	1	1	0	1	1	0	1	1	1

- Las 2 unidades se convierten en 0 a cambio de convertir las 3 "decenas" binarias en 4.
- Las 4 "decenas" binarias se convierten en 0 a cambio de convertir los 4 "millares" en 5.
- Las 3 "centenas" binarias se convierten en 1 a cambio de convertir los 5 "millares" en 6.
- Los $6 = 110_2$ "millares" se convierten en 0 a cambio de sumar una "decena de millar" y una "centena de millar", etc.

En el proceso aparecen dos dígitos 1 "fuera de rango", que podríamos haber puesto directamente como una unidad y una "decena", y así obtenemos el resto módulo M_{17} de s_{11}.

En la primera tabla vemos que $s_{14} + 2^9 \equiv 2^{17} - 1 = M_{17} \equiv 0 \pmod{M_{17}}$, de modo que $s_{14} \equiv -2^9 \pmod{M_{17}}$, luego

$$s_{14}^2 \equiv 2^{18} \equiv 2 \pmod{M_{17}},$$

luego $s_{15} \equiv 0 \pmod{M_{17}}$ y esto prueba que M_{17} es primo.

Lukas realizó un cálculo esencialmente en estos mismos términos[4] para comprobar que M_{127} es primo, para lo cual tuvo que trabajar con números de hasta 127 dígitos binarios, y calcular (módulo M_{127}) 125 términos de la sucesión s_n.

En 1883, el clérigo ruso Iván Pervushin aplicó el criterio de Lucas para probar que M_{61} es primo, y un matemático aficionado estadounidense llamado Ralph Ernest Powers comprobó que M_{89} y M_{107} son primos en 1911 y 1914, respectivamente. No fue hasta 1952 cuando el matemático estadounidense Raphael Mitchel Robinson, con la ayuda del SWAC (uno de los primeros ordenadores digitales que se construyeron) batió el récord de Lucas al comprobar (mediante el test de Lucas-Lehmer) que M_{521}, M_{607}, $M_{1\,279}$, $M_{2\,203}$ y $M_{2\,281}$ son primos (y también comprobó que todos los números de Mersenne M_p anteriores para valores de p distintos de

$$2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1\,279, 2\,203, 2\,281$$

son compuestos.) El número $M_{2\,281}$ tiene 687 cifras, pero sólo fue el mayor primo conocido durante apenas cinco años.

9.8 Sumas de Gauss generalizadas

Caracteres de Dirichlet Observemos que el carácter χ_k de un cuerpo cuadrático k es un carácter de Dirichlet en el sentido introducido en [ITAn 7.19].

En general, podemos definir un *carácter módulo m* como una aplicación

$$\chi : U_m \longrightarrow \mathbb{C} \setminus \{0\}$$

que cumpla la relación $\chi(xy) = \chi(x)\chi(y)$. Cualquier aplicación en estas condiciones define otra $\tilde{\chi} : \mathbb{Z} \longrightarrow \mathbb{C}$ mediante

$$\tilde{\chi}(n) = \begin{cases} \chi(\bar{n}) & \text{si } (m,n) = 1, \\ 0 & \text{si } (m,n) \neq 1, \end{cases}$$

y es inmediato que $\tilde{\chi}$ satisface la definición de carácter de Dirichlet dada en [ITAn 7.19]. Recíprocamente, cada carácter de Dirichlet $\tilde{\chi}$ en este sentido define un carácter χ módulo m mediante $\chi(\bar{n}) = \tilde{\chi}(n)$, de modo que no necesitamos distinguir χ de $\bar{\chi}$, y así podemos ver indistintamente a un carácter módulo m como una aplicación $\chi : \mathbb{Z} \longrightarrow \mathbb{C}$ o como $\chi : U_m \longrightarrow \mathbb{C}$.

Tal y como señalábamos en [ITAn], como $\chi(1) \neq 0$ y $\chi(1) = \chi(1 \cdot 1) = \chi(1)^2$, tiene que ser $\chi(1) = 1$. A su vez, si $c \in U_m$, se cumple que $c^{\phi(m)} = 1$, por lo que $\chi(c)^{\phi(m)} = \chi(1) = 1$, luego las imágenes de un carácter no son números complejos arbitrarios, sino que tienen que ser concretamente raíces de la unidad. Recordemos también que los caracteres módulo m forman un grupo con el producto dado por $(\chi\psi)(x) = \chi(x)\psi(x)$.

[4] Lucas partía de $s_0 = 3$ en lugar de $s_0 = 4$, lo cual sólo es válido para estudiar números M_p con $p \equiv -1$ (mód 4), mientras que Lehmer demostró que partiendo de $s_0 = 4$ el test es válido para cualquier primo p. Esta diferencia no altera la complejidad computacional.

Los caracteres de los cuerpos cuadráticos son lo que se conoce como caracteres de Dirichlet cuadráticos, es decir, caracteres que cumplen $\chi^2 = 1$ o, equivalentemente, que sólo toman los valores ± 1 (y el 0 cuando los consideramos con dominio en $\mathbb{Z}$).

Ejemplo: los caracteres módulo 8 Vamos a calcular todos los caracteres módulo 8. Tenemos que $U_8 = \{\bar{1}, \bar{3}, \bar{5}, \bar{7}\}$ tiene cuatro elementos, pero no hay raíces primitivas, sino que los cuatro elementos cumplen $x^2 = 1$. Esto se traduce en que todo carácter tiene que cumplir $\chi(x)^2 = 1$, luego $\chi(x) = \pm 1$. Teniendo en cuenta además que, necesariamente, $\chi(\bar{1}) = 1$, y que, como $\bar{7} = \bar{3} \cdot \bar{5}$, tiene que ser $\chi(\bar{7}) = \chi(\bar{3})\chi(\bar{5})$, tenemos sólo cuatro posibilidades:

	1	3	5	7
χ_0	1	1	1	1
δ	1	−1	1	−1
ϵ	1	−1	−1	1
$\delta\epsilon$	1	1	−1	−1

Observemos que

$$\chi_0(n) = \left(\frac{1}{n}\right), \qquad \delta(n) = \left(\frac{-1}{n}\right), \qquad \epsilon(n) = \left(\frac{2}{n}\right), \qquad (\delta\epsilon)(n) = \left(\frac{-2}{n}\right).$$

Estas representaciones en términos del símbolo de Legendre destacan un hecho que resulta ser relevante: los caracteres ϵ y $\delta\epsilon$ son "auténticos" caracteres módulo 8, en el sentido de que dependen del resto módulo 8 de los números sobre los que actúan, mientras que para calcular $\delta(n)$ no necesitamos realmente conocer el resto de n módulo 8, sino que nos basta con saber su resto módulo 4. En el caso de χ_0 no necesitamos saber nada de n o, equivalentemente, nos basta con su resto módulo 1. Esto nos lleva a la noción de carácter primitivo, que introducimos a continuación. ∎

Caracteres primitivos Observemos que si $m \mid m'$, todo carácter χ módulo m determina un carácter módulo m' dado por

$$\chi'(a) = \begin{cases} \chi(a) & \text{si } (a, m') = 1, \\ 0 & \text{si } (a, m') \neq 1. \end{cases}$$

Llamaremos a χ' el *carácter inducido* por χ. Cuando un carácter χ' módulo m' está inducido por otro carácter χ módulo m, sucede que el valor de $\chi'(a)$ depende en realidad del resto de a módulo m y no del resto módulo m'.

Por ejemplo, un carácter módulo 4 es el dado por $\delta(n) = (-1/n)$, el cual induce el carácter módulo 8 que también hemos llamado δ en el ejemplo anterior.

Otra forma conveniente de verlo es ésta: si $m \mid m'$, podemos considerar la aplicación

$$U_{m'} \longrightarrow U_m$$

dada por $\bar{a} \mapsto \bar{a}$.

Notemos que no depende del representante considerado en cada clase de restos, pues si $a \equiv a' \pmod{m'}$, también $a \equiv a' \pmod{m}$. Así, si χ es un carácter módulo m, el carácter que induce módulo m' es el que resulta de componer las aplicaciones:

$$U_{m'} \longrightarrow U_m \xrightarrow{\chi} \mathbb{C},$$

es decir, el que a cada $\bar{a} \in U_{m'}$ le asigna el resultado de pasar a U_m por la aplicación $\bar{a} \mapsto \bar{a}$ y luego aplicar χ.

Ejemplo La tabla siguiente muestra un carácter módulo 5 y el carácter que induce módulo 40:

U_{40}	1	11	21	31	7	17	27	37	3	13	23	33	9	19	29	39
U_5	1				2				3				4			
χ	1				i				$-i$				-1			

Observemos que, si llamamos χ' al carácter inducido, se cumple que $\chi'(2) = 0$, mientras que $\chi(2) = 0$. Pese a ello, si sólo tuviéramos la tabla de χ', podríamos calcular a partir de ella el valor de $\chi(2)$ sin más que observar que, en U_5, se cumple que $\bar{2} = \bar{7}$, por lo que $\chi(2) = \chi(7) = \chi'(7) = i$. ∎

En general, si $m \mid m'$, se cumple que la aplicación $U_{m'} \longrightarrow U_m$ es suprayectiva, es decir, que toda clase de U_m admite un representante primo con m', por lo que tiene antiimagen en $U_{m'}$, y en consecuencia siempre podemos reconstruir un carácter χ a partir de cualquier carácter χ' inducido por él.

Para probarlo observamos que si $\bar{a} \in U_m$, por el teorema chino del resto existe un a' que cumple $a' \equiv a \pmod{m}$ y $a' \equiv 1 \pmod{p}$ para todo primo p que divida a m' pero no a m. Entonces $(a', m') = 1$, por lo que $\bar{a}' \in U_{m'}$ tiene imagen $\bar{a}' = \bar{a} \in U_m$.

Esto implica que dos caracteres distintos en U_m no pueden inducir el mismo carácter en $U_{m'}$, pues ambos diferirán en una clase de U_m, luego los caracteres inducidos diferirán en las antiimágenes de dicha clase en $U_{m'}$.

Definición 9.18 Un carácter modular es *primitivo* si no está inducido por un carácter de módulo menor.

El teorema siguiente es útil para reconocer caracteres primitivos:

Teorema 9.19 *Un carácter χ módulo m es primitivo si y sólo si para todo divisor propio d de m existe un entero x tal que $(x, m) = 1$, $x \equiv 1 \pmod{d}$ y $\chi(x) \neq 1$.*

Demostración: Si χ no es primitivo está inducido por un carácter χ_0 módulo d, donde d es un divisor propio de m. Si $(x, m) = 1$ y $x \equiv 1 \pmod{d}$, entonces $\chi(x) = \chi_0(x) = \chi_0(1) = 1$.

Recíprocamente, si existe un divisor d de m tal que para todo $x \equiv 1 \pmod{d}$, $(x, m) = 1$ se cumple $\chi(x) = 1$, entonces si $x \equiv x' \pmod{d}$ y x, x' son primos con m, se cumple $\chi(x) = \chi(x')$.

En efecto, sea $\bar{x}'^{-1} = \bar{y}$, de manera que $x'y \equiv 1 \pmod{m}$, luego también $x'y \equiv 1 \pmod{d}$, luego $xy \equiv 1 \pmod{d}$, luego $\chi(\bar{x}\bar{x}'^{-1}) = \chi(\bar{x}\bar{y}) = 1$, luego $\chi(x) = \chi(x')$.

De aquí que podamos definir un carácter ψ módulo d mediante $\psi(a) = \chi(x)$, para cualquier x tal que $(x, m) = 1$ y $x \equiv a \pmod{d}$ (siempre existe tal x y acabamos de probar que $\chi(x)$ no depende de cuál elijamos). Claramente, el carácter ψ induce a χ. ∎

Nota La demostración de este teorema muestra que, en realidad, para que χ sea primitivo basta que cumpla la condición para todo divisor d de la forma $d = m/p$, donde p es un primo que divide a m, pues si χ está inducido por un carácter χ_0 módulo m_0, donde m_0 divide estrictamente a m, existe un primo p tal que $d_0 \mid m/p$, luego χ también está inducido por el carácter módulo $d' = m/p$ inducido por χ_0, y entonces d' no cumple la hipótesis. ∎

Vamos a usar este criterio (teniendo en cuenta la nota) para demostrar que los caracteres de los cuerpos cuadráticos son primitivos:

Teorema 9.20 *Los caracteres de los cuerpos cuadráticos son primitivos.*

Demostración: Sea $k = \mathbb{Q}(\sqrt{d})$ un cuerpo cuadrático, de modo que χ_k es un carácter módulo Δ_k. Tomemos un primo $p \mid \Delta_k$. Llamando $\Delta_0 = \Delta_k/p$, tenemos que encontrar un entero x tal que $(x, \Delta_k) = 1$, $x \equiv 1 \pmod{\Delta_0}$ y $\chi_k(x) = -1$.

Supongamos en primer lugar que $p \neq 2$. Entonces p divide a Δ con exponente 1, por lo que $(\Delta_0, p) = 1$. Tomamos entonces un entero s tal que $(s/p) = -1$. Por el teorema chino del resto existe un entero $x > 0$ que cumple

$$x \equiv s \pmod{p}, \qquad x \equiv 1 \pmod{4\Delta_0}.$$

Claramente $(x, \Delta_k) = 1$ y

$$\chi_k(x) = \left(\frac{\Delta_k}{x}\right) = \left(\frac{p}{x}\right)\left(\frac{\Delta_0}{x}\right) = \left(\frac{x}{p}\right)\left(\frac{x}{|\Delta_0|}\right) = \left(\frac{s}{p}\right) = -1,$$

donde hemos usado la ley de reciprocidad cuadrática para el símbolo de Jacobi. Notemos que si $\Delta_0 < 0$, entonces

$$\left(\frac{\Delta_0}{x}\right) = \left(\frac{-1}{x}\right)\left(\frac{|\Delta_0|}{x}\right) = \left(\frac{x}{|\Delta_0|}\right),$$

pues $x \equiv 1 \pmod{4}$.

Ahora supongamos que $p = 2$. Si $d \equiv -1 \pmod{4}$ y $\Delta_k = 4d$, tomamos $x > 0$ tal que $x \equiv -1 \pmod{4}$, $x \equiv 1 \pmod{d}$. Así tenemos que $(x, \Delta_k) = 1$ y también $x \equiv 1 \pmod{\Delta_0}$, pues $\Delta_0 = 2d$ y ciertamente $2 \mid x - 1$, $d \mid x - 1$.

Además:

$$\chi(x) = \left(\frac{\Delta_k}{x}\right) = \left(\frac{d}{x}\right) = -\left(\frac{x}{|d|}\right) = -\left(\frac{1}{|d|}\right) = -1,$$

donde, en el caso en que $d < 0$, usamos que $-d \equiv 1 \pmod 4$, por lo que

$$\left(\frac{d}{x}\right) = \left(\frac{-1}{x}\right)\left(\frac{|d|}{x}\right) = -\left(\frac{x}{|d|}\right).$$

Nos falta considerar el caso en que $d = 2d'$, de modo que $\Delta_k = 8d'$. Entonces tomamos $x \equiv 5 \pmod 8$, $x \equiv 1 \pmod{d'}$. Así $(x, \Delta_k) = 1$, $x \equiv 1 \pmod{\Delta_0}$, pues $\Delta_0 = 4d'$ y $x \equiv 1 \pmod 4$, $x \equiv 1 \pmod{d'}$. Además:

$$\chi(x) = \left(\frac{\Delta_k}{x}\right) = \left(\frac{2}{x}\right)\left(\frac{d'}{x}\right) = -\left(\frac{x}{|d'|}\right) = -\left(\frac{1}{|d'|}\right) = -1. \qquad \blacksquare$$

Puede probarse que los caracteres de los cuerpos cuadráticos son todos los caracteres cuadráticos primitivos.

Sumas de Gauss A cada carácter modular le podemos asignar una suma de Gauss:

Definición 9.21 Sea m un número natural y a un número entero, sea χ un carácter de Dirichlet módulo m y $\omega = e^{2\pi i/m}$. Se llama *suma de Gauss* de χ a la expresión[5]

$$G_a(\chi) = \sum_r \chi(r)\omega^{ar},$$

donde r recorre un conjunto completo de representantes de las clases de U_m. Escribiremos $G(\chi)$ en lugar de $G_1(\chi)$.

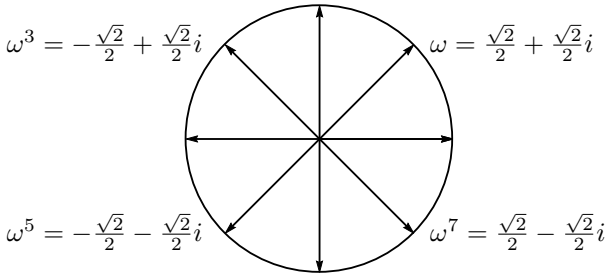
En realidad podemos definir sumas de Gauss sobre cualquier cuerpo k que contenga una raíz m-sima de la unidad ω de orden m, en cuyo caso ω^a recorre todas las raíces m-simas de la unidad, por lo que las sumas $G_a(\chi)$ son las sumas $G_1(\chi)$ que resultan de cambiar la elección de la raíz ω. Esto hace que, al considerar todos los valores posibles de a, la elección de la raíz se vuelve irrelevante a efectos puramente algebraicos. No obstante, aquí vamos a considerar únicamente sumas de Gauss en el cuerpo $\mathbb{C}$ y elegiremos siempre $\omega = e^{2\pi i/m}$.

En estos términos, la suma de Gauss $G(p)$ definida en la sección 7.3 es la suma $G_1(p)$ asociada al carácter $\chi(n) = (n/p)$.

Ejemplo Vamos a calcular las sumas de Gauss asociadas a los caracteres módulo 8 calculados en la página 319. Tenemos entonces que

$$\omega = \cos\frac{\pi}{4} + i\,\mathrm{sen}\,\frac{\pi}{4} = \frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}.$$

[5]En la sección 7.3 de [ITAl] definimos unas sumas de Gauss $G(p)$ que son, concretamente las sumas $G_1(\chi)$ que acabamos de definir, donde $\chi(n) = (n/p)$ es el carácter definido por el símbolo de Legendre.



Es fácil ver entonces que

$$\begin{aligned} G(\chi_0) &= \omega + \omega^3 + \omega^5 + \omega^7 = 0,\\ G(\delta) &= \omega - \omega^3 + \omega^5 - \omega^7 = 0,\\ G(\epsilon) &= \omega - \omega^3 - \omega^5 + \omega^7 = \sqrt{8},\\ G(\delta\epsilon) &= \omega + \omega^3 - \omega^5 - \omega^7 = \sqrt{8}i. \end{aligned}$$

Enseguida veremos que las sumas nulas se deben a que los caracteres correspondientes no son primitivos. Por ejemplo, si consideramos a δ como carácter módulo 4, entonces

$$G(\delta) = i - i^3 = 2i. \qquad \blacksquare$$

El teorema siguiente es un primer ejemplo de la relevancia que tiene que un carácter sea primitivo:

Teorema 9.22 *Sea χ un carácter primitivo. Entonces*

$$G_a(\chi) = \overline{\chi(a)}\, G(\chi).$$

DEMOSTRACIÓN: Sea $d = (a, m)$ y sea $m = td$. Entonces ω^a tiene orden t y $\omega^{au} = \omega^a$ siempre que $u \equiv 1 \pmod t$. Si $d \neq 1$ entonces t es un divisor propio de m y por el teorema 9.19 existe un entero u tal que $u \equiv 1 \pmod t$, $(u, m) = 1$ y $\chi(u) \neq 1$.

Cuando r recorre un conjunto completo de representantes de las clases de U_m lo mismo le sucede a ur, luego

$$G_a(\chi) = \sum_r \chi(ur)\omega^{aur} = \chi(u) \sum_r \chi(r)\omega^{ar} = \chi(u)G_a(\chi).$$

Puesto que $\chi(u) \neq 1$ ha de ser $G_a(\chi) = 0$. Por otro lado, $\overline{\chi(a)} = 0$, luego se cumple la igualdad.

Por el contrario, si $(a, m) = 1$, cuando r recorre un conjunto completo de representantes de las clases de U_m lo mismo le sucede a ar, luego

$$\chi(a)G_a(\chi) = \sum_r \chi(ar)\omega^{ar} = \sum_r \chi(r)\omega^r = G(\chi),$$

y multiplicando por $\overline{\chi(a)} = \chi(a)^{-1}$ obtenemos la igualdad. $\blacksquare$

La fórmula (7.1) admite esta generalización sencilla:

Teorema 9.23 *Si χ es un carácter primitivo módulo m, entonces*

$$|G_a(\chi)| = \sqrt{m}.$$

DEMOSTRACIÓN: Vamos a calcular de dos formas distintas la suma

$$S = \sum_{a=0}^{m-1} G_a(\chi)\overline{G_a(\chi)}.$$

Por el teorema anterior es

$$S = \sum_{a=0}^{m-1} \overline{\chi(a)}G(\chi)\chi(a)\overline{G(\chi)} = \sum_{a=0}^{m-1} |\chi(a)|^2|G(\chi)|^2 = \phi(m)|G(\chi)|^2.$$

En cambio, si tenemos en cuenta la definición de las sumas de Gauss:

$$S = \sum_{a=0}^{m-1}\sum_{s,t} \chi(s)\chi(t)\omega^{a(s-t)} = \sum_{s,t}\chi(s)\chi(t)\sum_{a=0}^{m-1}(\omega^{s-t})^a,$$

donde s, t recorren un conjunto de representantes de las clases de U_m. La última suma vale m si $s = t$ y en caso contrario vale

$$\frac{(\omega^{s-t})^m - 1}{\omega^{s-t} - 1} = 0,$$

pues $\omega^m = 1$. Por lo tanto:

$$S = \sum_s \chi(s)^2 m = \phi(m)m.$$

Al comparar las dos expresiones resulta que $|G(\chi)|^2 = m$. ∎

No se conoce ninguna expresión explícita para $G(\chi)$ cuando χ es un carácter arbitrario, pero para el caso de los caracteres cuadráticos el teorema 7.10 admite la generalización siguiente:

Teorema 9.24 *Si χ es un carácter cuadrático primitivo módulo m, entonces*

$$G(\chi) = \begin{cases} \sqrt{m} & \text{si } \chi(-1) = 1, \\ i\sqrt{m} & \text{si } \chi(-1) = -1. \end{cases}$$

Hemos expresado así el teorema porque es su enunciado más natural, pero aquí probaremos un enunciado equivalente. Como ya hemos señalado —aunque no lo hemos demostrado— los caracteres cuadráticos primitivos son exactamente los caracteres de los cuerpos cuadráticos, así que vamos a demostrar que si χ_k es el carácter del cuerpo cuadrático k, entonces

$$G(\chi_k) = \begin{cases} \sqrt{\Delta_k} & \text{si } \Delta_k > 0, \\ i\sqrt{|\Delta_k|} & \text{si } \Delta_k < 0. \end{cases} \tag{9.7}$$

(Recordemos la propiedad 4 de la página 304). Éste es el resultado que realmente utilizaremos más adelante.

Demostraremos esta relación reduciéndola en última instancia a 7.10 a través del teorema siguiente:

Teorema 9.25 *Sean $\chi_1, \dots, \chi_n$ caracteres módulo $m_1, \dots, m_n$ respectivamente, donde los números m_i son primos entre sí dos a dos. Sea $m = m_1 \cdots m_n$, sea χ_i^* el carácter móduño m inducido por χ_i y sea $\chi = \chi_1^* \cdots \chi_n^*$. Entonces*

$$G_a(\chi) = G_a(\chi_1) \cdots G_a(\chi_n)\, \chi_1(m/m_1) \cdots \chi_n(m/m_n).$$

DEMOSTRACIÓN: Supongamos que el teorema se cumple para $n = 2$, es decir, que

$$G_a(\chi_1^*\chi_2^*) = G_a(\chi_1)G_a(\chi_2)\, \chi_1(m_2)\chi_2(m_1).$$

Entonces una simple inducción sobre n prueba el caso general, pues si el teorema es cierto para $n-1$, entonces

$$G_a(\chi_1^* \cdots \chi_n^*) = G_a(\chi_1^* \cdots \chi_{n-1}^*)\, G_a(\chi_n)(\chi_1^* \cdots \chi_{n-1}^*)(m_n)\, \chi_n(\frac{m}{m_n})$$

$$= G_a(\chi_1) \cdots G_a(\chi_n)\chi_1(\frac{m}{m_n m_1}) \cdots \chi_{n-1}(\frac{m}{m m_{n-1}})\chi_1(m_n) \cdots \chi_{n-1}(m_n)\chi_n(\frac{m}{m_n})$$

$$= G_a(\chi_1) \cdots G_a(\chi_n)\, \chi_1(m/m_1) \cdots \chi_n(m/m_n).$$

Centrándonos, pues, en el caso $n = 2$, observamos que la aplicación

$$U_{m_1} \times U_{m_2} \longrightarrow U_m$$

definida como $\big([u],[v]\big) \mapsto [um_2 + vm_1]$ es biyectiva. En efecto, es claro que está bien definida (es decir, que la imagen de un par de clases no depende de los representantes elegidos) y además la relación de Bezout nos permite expresar $1 = xm_1 + ym_2$, luego todo número entero n puede expresarse en la forma $n = ynm_2 + xnm_1$, luego está en la imagen de la aplicación, y como ambos conjuntos tienen $\phi(m) = \phi(m_1)\phi(m_2)$ elementos, cada clase de U_m no puede tener más que una antiimagen en el producto.

Además, si $\omega = \cos(2\pi/m) + i\,\mathrm{sen}(2\pi/m)$, entonces

$$\omega^{m_2} = \cos(2\pi/m_1) + i\,\mathrm{sen}(2\pi/m_1) \quad \text{y} \quad \omega^{m_1} = \cos(2\pi/m_2) + i\,\mathrm{sen}(2\pi/m_2).$$

Por lo tanto,

$$G_a(\chi_1)G_a(\chi_2)\, \chi_1(m_2)\chi_2(m_1) = \Big(\sum_{u,v} \chi_1(u)\chi_2(v)\omega^{m_2 au + m_1 av}\Big)\chi_1(m_2)\chi_2(m_1)$$

$$= \sum_{u,v} \chi_1(m_2u)\chi_2(m_1v)\omega^{a(m_2u+m_1v)}$$

$$= \sum_{u,v} \chi_1(m_2u + m_1v)\chi_2(m_2u + m_1v)\omega^{a(m_2u+m_1v)} = \sum_r \chi_1(r)\chi_2(r)\omega^{ar}$$

$$= \sum_r \chi(r)\omega^{ar} = G_a(\chi),$$

donde u varía en U_{m_1}, v varía en U_{m_2} y r en U_m. ∎

Ahora ya podemos probar (9.7). Sea $k = \mathbb{Q}\big(\sqrt{d}\,\big)$ y consideremos en primer lugar el caso en que $d \equiv 1$ (mód 4), con lo que $\Delta_k = d$.

Pongamos que $d = p_1 \cdots p_r$ y sea $p_i^* = \pm p_i$ con el signo adecuado para que $p_i^* \equiv 1 \pmod 4$. Entonces $d = \pm p_1^* \cdots p_r^*$, pero al tomar restos módulo 4 vemos que el signo tiene que ser positivo, es decir, que $d = p_1^* \cdots p_r^*$. Entonces

$$\chi_k(n) = \left(\frac{\Delta_k}{n}\right) = \left(\frac{p_1^*}{n}\right) \cdots \left(\frac{p_r^*}{n}\right) = \left(\frac{n}{p_1}\right) \cdots \left(\frac{n}{p_r}\right).$$

En efecto, si $p_i^* > 0$ es inmediato por la ley de reciprocidad cuadrática para el símbolo de Jacobi, y si $p_i^* < 0$ entonces $p_i \equiv -1 \pmod 4$, luego

$$\left(\frac{p_i^*}{n}\right) = \left(\frac{-1}{n}\right)\left(\frac{p_i}{n}\right) = \left(\frac{n}{p_i}\right),$$

porque para aplicar la ley de reciprocidad cuadrática hay que añadir precisamente el signo $(-1/n)$. Esto significa que $\chi_k = \chi_1^* \cdots \chi_r^*$, donde $\chi_i(n) = (n/p_i)$. Por el teorema anterior:

$$G(\chi_k) = G(\chi_1) \cdots G(\chi_r)\chi_1(|\Delta_k|/p_1) \cdots \chi_r(|\Delta_k|/p_r).$$

Sea s el número de primos $p_i \equiv -1 \pmod 4$. Entonces, el teorema 7.10 nos da que

$$G(\chi_k) = i^s\sqrt{p_1} \cdots \sqrt{p_r}\left(\frac{|d|/p_1}{p_1}\right) \cdots \left(\frac{|d|/p_r}{p_r}\right) = i^s\sqrt{|\Delta_k|}\prod_{i\neq j}\left(\frac{p_i}{p_j}\right)\left(\frac{p_j}{p_i}\right).$$

Por la ley de reciprocidad cuadrática, los factores iguales a -1 son los correspondientes a los $s(s-1)$ pares de primos $p_i \equiv -1 \pmod 4$, luego

$$G(\chi_k) = i^s\sqrt{|\Delta_k|}(-1)^{s(s-1)/2} = i^{s+s^2-s}\sqrt{|\Delta_k|} = i^{s^2}\sqrt{|\Delta_k|},$$

pero si s es par, entonces $s^2 \equiv 0 \pmod 4$, luego $i^{s^2} = 1$ y $\Delta_k > 0$, mientras que si s es impar, entonces $s^2 \equiv 1 \pmod 4$, luego $i^{s^2} = i$ y $\Delta_k < 0$, luego la conclusión es precisamente (9.7).

Supongamos ahora que $d \equiv -1 \pmod 4$, con lo que $\Delta_k = 4d$. Entonces, el cuerpo cuadrático $\mathbb{Q}\big(\sqrt{-d}\,\big)$ tiene discriminante $\Delta^* = -d \equiv 1 \pmod 4$, y ya tenemos probado que su carácter χ^* cumple lo requerido. Además $\Delta_k = -4\Delta^*$ y

$$\chi_k(n) = \left(\frac{\Delta_k}{n}\right) = \left(\frac{-4\Delta^*}{n}\right) = \left(\frac{-1}{n}\right)\left(\frac{\Delta^*}{n}\right) = (\chi_1\chi_2)(n),$$

donde $\chi_1(n) = (-1/n)$ y χ_2 es el carácter del cuerpo $\mathbb{Q}\big(\sqrt{\Delta^*}\,\big)$, para el que ya tenemos probado el teorema. El teorema anterior nos da que

$$G(\chi_k) = G(\chi_1)G(\chi_2)\chi_1(|\Delta^*|)\chi_2(4) = 2iG(\chi_2)\chi_1(|\Delta^*|),$$

donde hemos usado que ya habíamos calculado $G(\chi_1) = 2i$. Si $\Delta_k > 0$, entonces $\Delta^* < 0$ y $|\Delta^*| \equiv -1 \pmod 4$, luego

$$G(\chi) = 2i \cdot i\sqrt{|\Delta^*|}(-1) = \sqrt{\Delta_k}.$$

En cambio, si $\Delta_k < 0$, entonces

$$G(\chi) = 2i\sqrt{\Delta^*} \cdot 1 = i\sqrt{|\Delta_k|},$$

como había que probar.

Supongamos finalmente que d es par, digamos $d = \pm 2\Delta^*$, donde el signo lo elegimos de modo que $\Delta^* \equiv 1 \pmod 4$. Entonces $\Delta_k = \pm 8\Delta^*$ y

$$\chi_k(n) = \left(\frac{\pm 8\Delta^*}{n}\right) = \left(\frac{\pm 2}{n}\right)\left(\frac{\Delta^*}{n}\right) = (\chi_1\chi_2)(n),$$

donde $\chi_1(n) = (\pm 2/n)$ y χ_2 es el carácter del cuerpo $\mathbb{Q}\big(\sqrt{\Delta^*}\big)$, para el que ya tenemos probado el teorema. El teorema anterior nos da entonces que

$$\begin{aligned} G(\chi_k) &= G(\chi_1)G(\chi_2)\chi_1(|\Delta^*|)\chi_2(8) = G(\chi_1)G(\chi_2)\left(\frac{\pm 2}{|\Delta^*|}\right)\left(\frac{2}{|\Delta^*|}\right) \\ &= G(\chi_1)G(\chi_2)\left(\frac{\pm 1}{|\Delta^*|}\right). \end{aligned}$$

Sabemos que

$$G(\chi_1) = \begin{cases} \sqrt{8} & \text{si } \Delta_k/\Delta^* > 0, \\ \sqrt{8}i & \text{si } \Delta_k/\Delta^* < 0, \end{cases} \qquad G(\chi_2) = \begin{cases} \sqrt{\Delta^*} & \text{si } \Delta^* > 0, \\ \sqrt{|\Delta^*|}i & \text{si } \Delta^* < 0. \end{cases}$$

Distinguiendo los cuatro casos posibles según los signos de Δ_k y Δ^* obtenemos que $G(\chi_k)$ cumple siempre (9.7). ∎

Nota De la demostración anterior se desprende un hecho que necesitaremos más adelante: si Δ_k es par, entonces

$$\chi_k(n + \Delta_k/2) = -\chi_k(n).$$

En efecto, en la prueba hemos visto que $\chi_k = \chi_1\chi_2$, donde $\chi_1(n) = (-1/n)$ cuando $\Delta_k = -4\Delta^*$ o $\chi_1(n) = (\pm 2/n)$ cuando $\Delta_k = \pm 8\Delta^*$ y χ_2 es el carácter del cuerpo $\mathbb{Q}\big(\sqrt{\Delta^*}\big)$.

En ambos casos $\Delta_k/2$ es múltiplo de Δ^*, luego $\chi_2(n+\Delta_k/2) = \chi_2(n)$. Basta probar que $\chi_1(n + \Delta_k/2) = -\chi_1(n)$.

Si $\Delta_k = -4\Delta^*$, con $\Delta^* \equiv 1 \pmod 4$, entonces $\Delta_k/2 = -2\Delta^* \equiv 2 \pmod 4$, mientras que si $\Delta_k = \pm 8\Delta^*$, con $\Delta^* \equiv 1 \pmod 4$, entonces $\Delta^* \equiv 1, 5 \pmod 8$ y en ambos casos $\Delta_k/2 = \pm 4\Delta^* \equiv 4 \pmod 8$, y todo se reduce a comprobar que $\chi_1(n+2) = -\chi_1(n)$ en el primer caso y $\chi_1(n+4) = -\chi_1(n)$ en el segundo caso.

Concretamente, llamando δ, ϵ y $\delta\epsilon$ a los tres caracteres posibles (véase el ejemplo de la página 319), en el primer caso basta observar que $\delta(3) = -\delta(1)$, y en el segundo

$$\epsilon(5) = -\epsilon(1), \quad \epsilon(7) = -\epsilon(3), \quad (\delta\epsilon)(5) = -(\delta\epsilon)(1), \quad (\delta\epsilon)(7) = -(\delta\epsilon)(3).$$

∎

Capítulo X

Fracciones continuas

Henry Dudeney fue un matemático británico famoso por sus acertijos matemáticos. En una de sus colecciones de "diversiones matemáticas" (publicado en 1917) afirmó que el rey Harold II de Inglaterra, que en 1066 se enfrentó al ejército normando de Guillermo el Conquistador en la batalla de Hastings, contaba con 61 divisiones de soldados que formaban dispuestas en cuadrados idénticos, pero que formaban uno solo (un único cuadrado) cuando se unía a ellos su general. El acertijo consistía en averiguar cuántos soldados tenía Harold.[1] Equivalentemente, se trata de encontrar las soluciones enteras de la ecuación $61y^2 + 1 = x^2$ o, equivalentemente:

$$x^2 - 61y^2 = 1.$$

Se trata de un ejemplo de ecuación de Pell, y en este capítulo veremos cómo resolver cualquiera de ellas. La ecuación de Pell está relacionada con las unidades de los cuerpos cuadráticos reales, como ilustra el ejemplo siguiente:

Números triangulares cuadrados Vamos a abordar el problema siguiente:

Encontrar los números que son a la vez triangulares y cuadrados.

Es fácil ver que un número m es triangular si y solo si $8m+1$ es cuadrado, así que se trata de encontrar todos los números naturales (no nulos) y^2 que cumplen $8y^2 + 1 = x^2$, para cierto número natural x. Por lo tanto, el problema equivale a encontrar las soluciones naturales de la ecuación de Pell $x^2 - 8y^2 = 1$, pues es obvio que cualquier solución cumplirá que $x = 2m + 1$ es impar. A su vez, esto equivale a encontrar las soluciones de la ecuación $x^2 - 2y^2 = 1$ con $x = 2m + 1$ impar e $y = 2n$ par. Pero esto último equivale a encontrar las unidades de, anillo $\mathbb{Z}\big[\sqrt{2}\,\big]$ de la forma

$$\epsilon = (2m + 1) + 2n\sqrt{2}.$$

[1] El lector debe tener presente que — a pesar de la ambientación histórica del problema— la respuesta no es nada realista, por eso en algunas versiones del problema el número de divisiones se reduce a 13.

(En realidad queremos también que $n > 0$.) Sabemos que la unidad fundamental de este anillo es $\eta = 1+\sqrt{2}$, luego las soluciones que buscamos serán de la forma $\epsilon = x+y\sqrt{2} = \eta^k$, para los exponentes k que hagan que y sea par (notemos que tiene que ser $k > 0$, pues si $k < 0$ obtenemos los conjugados de las soluciones con $k > 0$, que tienen la y negativa). La tabla siguiente muestra las primeras potencias de η:

k	1	2	3	4	5	6
η^k	$1+\sqrt{2}$	$3+2\sqrt{2}$	$7+5\sqrt{2}$	$17+12\sqrt{2}$	$41+29\sqrt{2}$	$99+70\sqrt{2}$
m	–	1	–	8	–	49
n	–	1	–	6	–	35

De aquí obtenemos los números triangulares cuadrados 1^2, 6^2 y 35^2. Teniendo en cuenta que

$$(x+y\sqrt{2})(1+\sqrt{2}) = (x+2y)+(x+y)\sqrt{2},$$

es fácil ver que η^k tiene el coeficiente de $\sqrt{2}$ par si y sólo si k es par, por lo que las soluciones que buscamos vienen determinadas por las potencias de $\eta^2 = 3+2\sqrt{2}$. Explícitamente:

$$(x+y\sqrt{2})(3+2\sqrt{2}) = (3x+4y)+(2x+3y)\sqrt{2},$$

por lo que si tenemos unos valores de m y n que satisfacen la ecuación de partida, correspondientes a $x = 2m+1$, $y = 2n$, los siguientes corresponderán a

$$x' = 3(2m+1)+8n = 6m+8n+3, \qquad y' = 2(2m+1)+6n = 4m+6n+2,$$

que a su vez corresponden a

$$m' = \frac{6m+8n+3-1}{2} = 3m+4n+1, \qquad n' = \frac{4m+6n+1}{2} = 2m+3n+1.$$

La conclusión es que los pares (m,n) tales que el m-simo número triangular coincide con el cuadrado n-simo pueden obtenerse recurrentemente partiendo de $(m_0,n_0) = (1,1)$ mediante la relación

$$(m_{i+1},n_{i+1}) = (3m_i+4n_i+1, 2m_i+3n_i+1).$$

∎

Ejercicio: Encontrar con la ayuda de un ordenador (mediante una búsqueda) el menor número de soldados que podría tener el ejército del rey Harold si se cumplieran las condiciones indicadas por Dudeney.

En el capítulo anterior probamos la existencia de unidades fundamentales de los cuerpos cuadráticos a partir de un teorema de Dirichlet de aproximación diofántica. Sin embargo, el argumento no es nada eficiente a la hora de encontrar soluciones explícitas. Aquí vamos a proporcionar una técnica de aproximación diofántica mucho más eficiente que nos permitirá resolver cualquier ecuación de Pell y calcular la unidad fundamental de cualquier cuerpo cuadrático real. Antes planteamos un problema que puede resolverse con dicha técnica, por si el lector quiere pensarlo previamente:

Ejemplo *En una clase de menos de* 100 *alumnos, el* 67.19% *ha aprobado un examen (el porcentaje ha sido redondeado al decimal exacto más próximo con dos cifras decimales). Determinar cuántos alumnos han aprobado y cuántos han suspendido.*

10.1 Desarrollos en fracción continua

Sabemos —y es bien sabido— que todo número real, como

$$\pi = 3.141592653\ldots$$

puede aproximarse con precisión arbitraria por números racionales, por ejemplo, por los que resultan de truncar su desarrollo decimal (o en cualquier otra base):

$$3,\quad 3.1,\quad 3.14,\quad 3.141,\quad 3.1415,\quad 3.14159,\ldots$$

Si retenemos n cifras decimales obtenemos una aproximación con un error de a lo sumo 10^{-n}. Esto supone aproximar un número real α mediante un número racional p/q (donde $q = 10^n$) de modo que

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q}.$$

Sin embargo, el teorema de Dirichlet 9.9 nos dice que, para todo $M > 0$, podemos aproximar α mediante un número racional p/q tal que $0 < q < M$ y

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{Mq} < \frac{1}{q^2}.$$

Vamos a ver cómo encontrar fácilmente aproximaciones diofánticas que cumplan esta condición.

Definición 10.1 Partamos de una sucesión de números enteros $a_0, a_1, a_2, \ldots$, todos positivos salvo quizá el primero. Llamaremos

$$\begin{aligned}
[a_0] &= a_0,\\
[a_0, a_1] &= a_0 + \frac{1}{a_1},\\
[a_0, a_1, a_2] &= a_0 + \frac{1}{a_1 + \frac{1}{a_2}},\\
[a_0, a_1, a_2, a_3] &= a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3}}},
\end{aligned}$$

etc.

En general tenemos definido el número racional $[a_0, \ldots, a_n]$ para todo n, que es no nulo si $n \geq 1$. Una definición formal se da por recurrencia de derecha a izquierda, es decir:

$$x_0 = a_n,\quad x_{i+1} = a_{n-(i+1)} + 1/x_i,\quad [a_0, \ldots, a_n] = x_n.$$

Ejemplo Para calcular

$$[5,1,3,3,2] = 5 + \frac{1}{1+\frac{1}{3+\frac{1}{3+\frac{1}{2}}}} = \frac{173}{30}$$

podemos empezar por el 2:

$$2 \mapsto 3 + \frac{1}{2} = \frac{7}{2} \mapsto 3 + \frac{1}{7/2} = \frac{23}{7} \mapsto 1 + \frac{1}{23/7} = \frac{30}{23} \mapsto 5 + \frac{1}{30/23} = \frac{173}{30}.$$

■

Las expresiones $[a_0, \ldots, a_n]$ se llaman *fracciones continuas* (finitas).[2]

A continuación vamos a ver que este método no es el más conveniente para calcular fracciones continuas, sino que es mucho más práctico calcular las fracciones sucesivas

$$[5], \quad [5,1], \quad [5,1,3], \quad [5,1,3,3], \quad [5,1,3,3,2],$$

dado que existe una relación recurrente muy simple para calcular cada una de ellas a partir de las dos anteriores.

En general, llamaremos $r_n = [a_0, \ldots, a_n] = p_n/q_n$, donde p_n y q_n son enteros primos entre sí $q_n > 0$ (convenimos que si $a_0 = 0$, entonces $p_0 = 0$, $q_0 = 1$).

Teorema 10.2 *Con la notación anterior:*

$$p_0 = a_0, \quad q_0 = 1, \quad p_1 = a_0a_1 + 1, \quad q_1 = a_1,$$

$$p_n = a_np_{n-1} + p_{n-2}, \qquad q_n = a_nq_{n-1} + q_{n-2}.$$

Demostración: Los casos $n = 0, 1, 2$ se comprueban directamente. Hay que probar que los valores dados por las fórmulas (en estos tres casos) son realmente primos entre sí, pero esto se ve fácilmente por los métodos usuales.

Supongámoslo cierto para $n - 1 \geq 2$ y probémoslo para n. Definimos los enteros primos entre sí

$$\frac{p'_j}{q'_j} = [a_1, \ldots, a_{j+1}], \quad j = 0, 1, 2, \ldots$$

Por la hipótesis de inducción aplicada a $n - 1$ se cumplen las fórmulas

$$p'_{n-1} = a_np'_{n-2} + p'_{n-3}, \qquad q'_{n-1} = a_nq'_{n-2} + q'_{n-3}. \tag{10.1}$$

Por otra parte $\dfrac{p_j}{q_j} = a_0 + \dfrac{q'_{j-1}}{p'_{j-1}}$, luego

$$p_j = a_0p'_{j-1} + q'_{j-1}, \qquad q_j = p'_{j-1}, \tag{10.2}$$

[2]Observemos que el adjetivo "continua" no debe entenderse en relación con ninguna clase de "continuidad", sino en el sentido de "fracciones repetidas".

donde hemos usado que si $(p'_{j-1}, q'_{j-1}) = 1$, los valores que dan estas fórmulas también son primos entre sí. Haciendo $j = n$ en (10.2) y usando (10.1) obtenemos

$$\begin{aligned} p_n &= a_0(a_n p'_{n-2} + p'_{n-3}) + (a_n q'_{n-2} + q'_{n-3}) \\ &= a_n(a_0 p'_{n-2} + q'_{n-2}) + a_0 p'_{n-3} + q'_{n-3}, \\ q_n &= a_n q'_{n-2} + q'_{n-3}. \end{aligned}$$

Aplicando (10.2) con $j = n-1$ y $n-2$ se deduce

$$p_n = a_n p_{n-1} + p_{n-2}, \qquad q_n = a_n q_{n-1} + q_{n-2}. \qquad \blacksquare$$

En la práctica, en lugar de recordar la definición de p_1 y q_1, es útil observar que éstos pueden calcularse con la fórmula general para p_n y q_n si convenimos en definir $p_{-1} = 1$, $q_{-1} = 0$.

Ejemplo Calculemos de nuevo

$$[5,1,3,3,2] = 5 + \frac{1}{1 + \frac{1}{3 + \frac{1}{3 + \frac{1}{2}}}} = \frac{173}{30}.$$

n	-1	0	1	2	3	4
a_n	$-$	5	1	3	3	4
p_n	1	5	6	23	75	173
q_n	0	1	1	4	13	30

Observemos que

$$r_0 = \frac{5}{1} < r_2 = \frac{23}{4} < r_4 = \frac{173}{30} < r_3 = \frac{75}{13} < r_1 = \frac{6}{1} \qquad \blacksquare$$

Ejercicio: Calcular $[4,1,8,30,1,3]$ y ordenar los números $r_0, \ldots, r_5$.

La ordenación que estamos observando en los números racionales r_i que resultan de ir prolongando una fracción continua no es casual, sino que es una consecuencia sencilla del teorema anterior. Éste muestra claramente que las sucesiones p_n y q_n son estrictamente crecientes y además:

Teorema 10.3 *Con la notación anterior,* $p_n q_{n+1} - p_{n+1} q_n = (-1)^{n+1}$ *o, lo que es lo mismo:* $r_n - r_{n+1} = (-1)^{n+1}/q_n q_{n+1}$.

Demostración: Claramente

$$\begin{aligned} p_n q_{n+1} - p_{n+1} q_n &= p_n(a_{n+1} q_n + q_{n-1}) - (a_{n+1} p_n + p_{n-1}) q_n \\ &= p_n q_{n-1} - p_{n-1} q_n = -(p_{n-1} q_n - p_n q_{n-1}), \end{aligned}$$

y como $p_0 q_1 - p_1 q_0 = a_0 a_1 - (a_0 a_1 + 1) = -1$, se cumple el teorema. $\blacksquare$

Esto implica que si tomamos una sucesión infinita $a_0, a_1, a_2, \ldots$ de números enteros positivos (salvo a lo sumo el primero, que puede ser negativo) y vamos calculando los números racionales $r_n = [a_0, \ldots, a_n]$, éstos estarán siempre dispuestos así:

$$r_0 < r_2 < r_4 < r_6 < \cdots < r_7 < r_5 < r_3 < r_1,$$

pues, por una parte,

$$r_{n+2} - r_n = r_{n+2} - r_{n+1} + r_{n+1} - r_n = (-1)^{n+1}/q_{n+1}q_{n+2} + (-1)^{n+1}/q_n q_{n+1},$$

luego la sucesión $\{r_{2n}\}$ es creciente y $\{r_{2m+1}\}$ es decreciente y, por otra parte, cualquier r_{2n} es menor que cualquier r_{2m+1}, pues, si por ejemplo $n \leq m$, tenemos que

$$r_{2n} \leq r_{2m} < r_{2m+1},$$

ya que el teorema anterior nos da que $r_{2m} - r_{2m+1} < 0$, y si es $m < n$, entonces

$$r_{2n} < r_{2n+1} < r_{2m+1}.$$

Ahora podemos aplicar la completitud de $\mathbb{R}$ para concluir el siguiente resultado fundamental:

Teorema 10.4 *Con la notación anterior, existe un único número real α tal que*

$$r_0 < r_2 < r_4 < r_6 < \cdots \alpha \cdots < r_7 < r_5 < r_3 < r_1.$$

Demostración: Basta considerar el conjunto A de todos los números reales que son menores que algún r_i y el conjunto B de los que son mayores o iguales que todos ellos. Es claro que todo elemento de A es menor que todo elemento de B, y que entre ambos conjuntos cubren todo $\mathbb{R}$. Además A no puede tener un máximo elemento, luego, por completitud, B tiene que tener un mínimo elemento α. Obviamente α cumple lo requerido, pero falta justificar que es el único.

Si hubiera dos, digamos $\alpha < \beta$, entonces, para todo n, se cumple que

$$0 < \beta - \alpha < r_{2n+1} - r_{2n} = \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}.$$

El hecho de que q_n sea una sucesión estrictamente creciente de números naturales implica claramente que $n \leq q_n$ (si el lector no lo reconoce como inmediato, puede probarlo fácilmente por inducción), por lo que

$$0 < \beta - \alpha < \frac{1}{n^2} < \frac{1}{n},$$

para todo $n > 0$, y esto contradice la propiedad arquimediana de $\mathbb{R}$. ∎

Ahora podemos extender el concepto de fracción continua:

Definición 10.5 La *fracción continua* asociada a la sucesión finita de números enteros $a_0, \dots, a_n$ (todos positivos salvo a lo sumo a_0) es el número racional $[a_0, \dots, a_n]$. La *fracción continua* asociada a una sucesión infinita $\{a_n\}_{n=0}^{\infty}$ en las mismas condiciones es el número real α dado por el teorema anterior, al que representaremos en la forma

$$[a_0, a_1, a_2, \dots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\dots}}}$$

Las fracciones continuas finitas $r_n = [a_0, \dots, a_n]$ se llaman *convergentes* de la fracción continua infinita $\alpha = [a_0, a_1, a_2, \dots]$.

Vamos a probar que, al igual que todo número real admite un desarrollo decimal, también admite un desarrollo en fracción continua, de modo que los convergentes serán aproximaciones racionales alternativas (y veremos que —en general— mucho mejores) a los desarrollos decimales truncados.

Para ello conviene observar que la definición 10.1 vale igualmente aunque los números a_i no sean enteros, simplemente bajo la hipótesis de que todos menos el primero sean positivos. Vamos a considerar únicamente expresiones de la forma $[a_0, a_1, \dots, a_{n-1}, x]$, donde todos los a_i son enteros no nulos (menos el primero) y $x > 0$ es un número real. Por ejemplo,

$$[a_0, x] = a_0 + \frac{1}{x}, \quad [a_0, a_1, x] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{x}}, \quad [a_0, a_1, a_2, x] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{x}}}.$$

Observemos que la función $[a_0, x]$ es decreciente (es menor cuanto mayor es x), lo que a su vez implica que $[a_0, a_1, x]$ es creciente y a su vez $[a_0, a_1, a_2, x]$ es decreciente, y así sucesivamente. En general, la función $[a_0, \dots, a_n, x]$ es creciente si n es impar y decreciente si n es par.

Tomemos un número real arbitrario, por ejemplo $\alpha = \pi = 3.141592653\dots$

Definimos $\alpha_0 = \alpha$ y $a_0 = E[\alpha]$ (en nuestro caso $a_0 = 3$). Si $\alpha = a_0$, entonces $\alpha = [a_0]$ ya es un desarrollo de α en fracción continua. En caso contrario $a_0 < \alpha < a_0 + 1$, luego $0 < \alpha - a_0 < 1$ y podemos definir $\alpha_1 = 1/(\alpha_0 - a_0) > 1$. En nuestro caso

$$\alpha_1 = \frac{1}{\pi - 3} = 7.062513306\dots$$

Además,

$$[a_0] = a_0 < \alpha = a_0 + \alpha - a_0 = a_0 + \frac{1}{\alpha_1} = [a_0, \alpha_1].$$

Definimos ahora $a_1 = E[\alpha_1] \geq 1$. Si $a_1 = \alpha_1$, ya tenemos el desarrollo en fracción continua $\alpha = [a_0, a_1]$. En caso contrario (como sucede en nuestro caso: $a_1 = 7$) tenemos que $a_1 < \alpha_1 < a_1 + 1$, luego $0 < \alpha_1 - a_1 < 1$ y podemos definir $\alpha_2 = 1/(\alpha_1 - a_1) > 1$. En nuestro caso

$$\alpha_2 = \frac{1}{\alpha_1 - 7} = 15.99659441\dots$$

Así

$$a_1 < \alpha_1 = a_1 + \alpha_1 - a_1 = a_1 + \frac{1}{\alpha_2},$$

luego, por la monotonía de $[a_0, a_1, x]$,

$$\alpha = [a_0, \alpha_1] = a_0 + \frac{1}{\alpha_1} = a_0 + \frac{1}{a_1 + \frac{1}{\alpha_2}} = [a_0, a_1, \alpha_2] < [a_0, a_1].$$

En total tenemos que

$$[a_0] < \alpha = [a_0, a_1, \alpha_2] < [a_0, a_1].$$

Nuevamente definimos $a_2 = E[\alpha_2] \geq 1$. Si $\alpha_2 = a_2$ obtenemos el desarrollo $\alpha = [a_0, a_1, a_2]$ y, en caso contrario tenemos que $a_2 < \alpha_2 < a_2 + 1$, luego $0 < \alpha_2 - a_2 < 1$ y podemos definir $\alpha_3 = 1/(\alpha_2 - a_2) > 1$. En nuestro caso

$$\alpha_3 = \frac{1}{\alpha_2 - 15} = 1.003417231\ldots$$

Así

$$a_2 < \alpha_2 = a_2 + \alpha_2 - a_2 = a_2 + \frac{1}{\alpha_3},$$

luego, por la monotonía de $[a_0, a_1, a_n, x]$,

$$\alpha = [a_0, a_1, \alpha_2] = a_0 + \frac{1}{a_1 + \frac{1}{\alpha_2}} = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\alpha_3}}} = [a_0, a_1, a_2, \alpha_3] > [a_0, a_1, a_2].$$

En total tenemos que

$$[a_0] < [a_0, a_1, a_2] < \alpha = [a_0, a_1, a_2, \alpha_2] < [a_0, a_1].$$

Es claro entonces que si, partiendo de $\alpha_0 = \alpha$, $a_0 = E[\alpha]$, vamos calculando

$$\alpha_{n+1} = E[1/(\alpha_n - a_n)], \qquad a_{n+1} = E[\alpha_{n+1}],$$

o bien llega un momento en el que $\alpha_n = a_n$, en cuyo caso $\alpha = [a_0, \ldots, a_n]$, o bien esto no sucede nunca, y entonces $\alpha = [a_0, \ldots, a_n, \alpha_n]$ para todo n, y α es mayor o menor que r_n según si n es par o impar, con lo que

$$r_0 < r_2 < r_4 < r_6 < \cdots \alpha \cdots < r_7 < r_5 < r_3 < r_1,$$

y esto significa, por definición, que $\alpha = [a_0, a_1, a_2, \ldots]$

Con esto hemos probado lo siguiente:

Teorema 10.6 *Todo número real α admite un desarrollo en fracción continua (que puede ser finito o infinito)*

$$\alpha = [a_0, a_1, a_2, \ldots]$$

Concretamente, los coeficientes a_i pueden calcularse junto a las "colas" α_n como $\alpha_0 = \alpha$, $a_0 = E[\alpha]$ y

$$\alpha_{n+1} = E[1/(\alpha_n - a_n)], \qquad a_{n+1} = E[\alpha_{n+1}],$$

de modo que la sucesión termina si $\alpha_n = a_n$. Además, $\alpha = [a_0, \ldots, a_n, \alpha_{n+1}]$.

La fracción continua de π Para $\alpha = \pi$ hemos obtenido los primeros términos del desarrollo:

$$\pi = [3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, \ldots]$$

Si el lector sabe que π es irracional, entonces sabe también que su desarrollo en fracción continua es necesariamente infinito, ya que las fracciones continuas finitas son números racionales.

Es interesante observar las aproximaciones racionales de π que proporciona su fracción continua:

n	-1	0	1	2	3	4
a_n	$-$	3	7	15	1	292
p_n	1	3	22	333	355	103 993
q_n	0	1	7	106	113	33 102

Vemos que los primeros convergentes son las aproximaciones que obtuvimos en el capítulo anterior:

$$\frac{22}{7} = 3.1428\ldots, \qquad \frac{333}{106} = 3.141509, \qquad \frac{355}{113} = 3.1415929\ldots$$

El siguiente es

$$\frac{103\,993}{33\,102} = 3.1415926530\ldots$$

que aproxima a π con error menor que una milmillonésima. En el capítulo anterior obtuvimos las tres aproximaciones anteriores por el procedimiento que usamos para probar el teorema 9.9, lo cual requiere cálculos muy laboriosos, mientras que ahora las hemos obtenido por un proceso que podemos llevar a cabo en unos segundos con una simple calculadora.

La fracción continua de e Por ejemplo, para calcular el desarrollo en fracción continua del número $e = 2.71828182845905\ldots$ sólo tenemos que ir calculando:

	α_n	a_n
0	2.71828182845905	2
1	1.39221119117733	1
2	2.54964677830384	2
3	1.81935024359808	1
4	1.22047928564544	1
5	4.53557347608674	4
6	1.86715743898796	1
7	2.15319312853624	1
8	6.52770793021197	6

(cada α_n resulta de restarle la parte entera al anterior y calcular el inverso) de modo que

$$e = [2, 1, 2, 1, 1, 4, 1, 1, 6, \ldots]$$

Ejercicio: Calcular los primeros convergentes del número e.

La fracción continua de $\sqrt{2}$ Observemos ahora que de $(\sqrt{2}-1)(\sqrt{2}+1) = 1$ se sigue que

$$\sqrt{2} = 1 + \frac{1}{1+\sqrt{2}}.$$

Si llamamos $\alpha = 1 + \sqrt{2}$, tenemos que $\alpha = 2 + \frac{1}{\alpha} = [2, \alpha]$. Por otra parte sabemos que $\alpha = [2, \alpha_1]$, luego tiene que ser $\alpha = \alpha_1$. Así pues, al calcular el desarrollo en fracción continua de α obtenemos siempre $\alpha_n = \alpha_0$ y, por lo tanto $a_n = a_0 = 2$. La conclusión es que

$$1 + \sqrt{2} = [2, 2, 2, \ldots]$$

luego

$$\sqrt{2} = [1, 2, 2, 2, \ldots]$$

Esto nos da una demostración alternativa de que $\sqrt{2}$ es irracional:

Teorema 10.7 *Las fracciones continuas infinitas son números irracionales.*

DEMOSTRACIÓN: Con la notación anterior, supongamos que una fracción continua infinita es un número racional $\alpha = p/q$ (con p y q primos entre sí).

Como la sucesión q_n es creciente, existe un n tal que $q < q_{n+1}$. Puesto que α está entre r_n y r_{n+1}, se cumple que $|\alpha - r_n| \le |r_n - r_{n+1}| = 1/q_n q_{n+1} < 1/q_n q$.

Pero por otro lado $|\alpha - r_n| = |p/q - p_n/q_n| = |pq_n - qp_n|/q_n q \ge 1/q_n q$, puesto que $p/q \ne p_n/q_n$, luego $|pq_n - qp_n| \ge 1$, contradicción. ∎

Fracciones continuas de números racionales Es interesante observar que cuando calculamos los coeficientes del desarrollo en fracción continua de un número racional hacemos los mismos cálculos que para aplicar el algoritmo de Euclides. En efecto, consideremos de nuevo el caso

	4 070
1	3 626
8	444
6	74
	0

que pusimos como ejemplo en la sección 2.1 y comparémoslo con el cálculo del desarrollo en fracción continua de $\alpha_0 = 4\,070/3\,626$. En el cálculo anterior, el 1 surge de que

$$4\,070 = 3\,626 \cdot 1 + 444,$$

pero esto implica que $a_0 = E[\alpha_0] = 1$. Además

$$\alpha_1 = \frac{1}{\frac{4\,070}{3\,626} - 1} = \frac{1}{\frac{444}{3\,626}} = \frac{3\,626}{444},$$

por lo que podemos volver a empezar, para concluir que, por los cálculos del algoritmo de Euclides, $a_1 = 8$, $a_2 = 6$ y en este punto termina el desarrollo, de modo que

$$\frac{4\,070}{3\,626} = [1, 8, 6] = 1 + \frac{1}{8 + \frac{1}{6}}.$$

Ahora bien, observemos ahora que

$$\frac{4\,070}{3\,626} = [1, 8, 6] = 1 + \frac{1}{8 + \frac{1}{6}} = 1 + \frac{1}{8 + \frac{1}{5 + \frac{1}{1}}} = [1, 8, 5, 1].$$

Esto es claramente un caso particular de un hecho general:

> *Toda fracción continua finita que no acabe en* 1 *se puede alargar restando* 1 *a su último término y añadiéndo un coeficiente* 1 *y, recíprocamente, toda fracción continua que termine en* 1 *se puede reducir sumando el último* 1 *al coeficiente precedente.*

Por lo tanto, cada número racional admite (al menos) dos desarrollos distintos en fracción continua. No sucede lo mismo con los números irracionales.

Teorema 10.8 *Todo número irracional admite un único desarrollo en fracción continua.*

Demostración: Supongamos que dos sucesiones definen la misma fracción continua $[a_0, a_1, \dots\,] = [b_0, b_1, \dots\,]$. Entonces $a_0 \leq [a_0, a_1, \dots\,] \leq a_0 + 1$ e igualmente con la otra fracción. Como ésta es irracional no se dan las igualdades, luego $a_0 = E\big([a_0, a_1, \dots\,]\big) = E\big([b_0, b_1, \dots\,]\big) = b_0$.

Ahora bien, es claro que

$$[a_0, a_1, \dots, a_n] = a_0 + \frac{1}{[a_1, \dots, a_n]}, \qquad [b_0, b_1, \dots, b_n] = b_0 + \frac{1}{[b_1, \dots, b_n]},$$

luego $[a_1, \dots, a_n] = [b_1, \dots, b_n]$, para todo n, de donde se sigue que

$$[a_1, a_2, \dots\,] = [b_1, b_2, \dots\,].$$

De aquí se sigue a su vez que $a_1 = b_1$ y, continuando así, llegamos a que todos los coeficientes coinciden. ∎

Una variante de la prueba del teorema anterior muestra que cada número racional admite únicamente dos desarrollos en fracción continua.

Aproximaciones diofánticas El teorema 10.3 afirma que $|r_n - r_{n+1}| = 1/q_n q_{n+1}$ para cualquier par de convergentes consecutivos de una fracción continua. Puesto que la fracción continua α se halla entre ambos, tenemos que

$$|\alpha - r_n| < 1/q_n q_{n+1} < 1/q_n^2.$$

Esto significa que los convergentes son buenas aproximaciones de sus límites. Podemos mejorar ligeramente este hecho observando que

$$|\alpha - r_n| + |\alpha - r_{n+1}| = |r_n - r_{n+1}| = 1/q_n q_{n+1}.$$

Cualquier par de números reales distintos cumple $xy < (x^2 + y^2)/2$, concluimos que

$$|\alpha - r_n| + |\alpha - r_{n+1}| < \frac{1}{2q_n^2} + \frac{1}{2q_{n+1}^2}.$$

Esto prueba que de cada dos convergentes consecutivos de un número irracional α, uno de ellos, p/q cumple $|\alpha - p/q| < 1/2q^2$. El resultado principal que necesitamos es el recíproco de este hecho.

Teorema 10.9 *Si α es un número irracional y p, q son naturales primos entre sí tales que $|\alpha - p/q| < 1/2q^2$, entonces p/q es un convergente de α.*

DEMOSTRACIÓN: Vamos a probar que si p y q son enteros cualesquiera tales que $0 < q < q_{n+1}$, entonces $|q\alpha - p| \geq |q_n\alpha - p_n|$. Esto significa que el convergente n-simo es la mejor aproximación racional de α con denominador menor que q_{n+1}.

En efecto, consideramos el sistema de ecuaciones

$$\begin{aligned} p &= up_n + vp_{n+1} \\ q &= uq_n + vq_{n+1} \end{aligned}$$

Por el teorema 10.3, se cumple que $p_n q_{n+1} - p_{n+1} q_n = (-1)^{n+1}$, de donde se sigue que el sistema tiene una solución entera, dada por

$$u = (-1)^{n+1}(q_{n+1}p - p_{n+1}q), \qquad v = (-1)^{n+1}(p_n q - q_n p).$$

Como $0 < q < q_{n+1}$, se ha de cumplir $u \neq 0$ y en el caso en que $v \neq 0$ entonces u y v tienen signos opuestos, y así

$$\begin{aligned} |q\alpha - p| &= |(uq_n + vq_{n+1})\alpha - (up_n + vp_{n+1})| \\ &= |u(q_n\alpha - p_n) + v(q_{n+1}\alpha - p_{n+1})| \geq |q_n\alpha - p_n|. \end{aligned}$$

Ahora, en las hipótesis del teorema, tomamos un n tal que $q_n \leq q < q_{n+1}$. Entonces

$$\left|\frac{p}{q} - \frac{p_n}{q_n}\right| \leq \left|\alpha - \frac{p}{q}\right| + \left|\alpha - \frac{p_n}{q_n}\right| = \frac{|\alpha q - p|}{q} + \frac{|\alpha q_n - p_n|}{q_n} \leq \left(\frac{1}{q} + \frac{1}{q_n}\right)|\alpha q - p|.$$

Como $q \geq q_n$ y $|\alpha q - p| < 1/2q$, concluimos que

$$\frac{|pq_n - qp_n|}{qq_n} < \frac{1}{qq_n},$$

y como el numerador es entero, ha de ser 0, luego p/q es el convergente n-simo. ∎

El teorema anterior vale también si $\alpha = r/s > 0$ es un número racional (donde $(r,s) = 1$) si suponemos además que $1 \le q < s$, pues así s es el denominador del último convergente de α, luego existe igualmente un n tal que $q_n \le q < q_{n+1}$, y toda la prueba vale igualmente. Esta observación nos permite resolver el problema que habíamos planteado justo antes de esta sección:

Ejemplo *En una clase de menos de* 100 *alumnos, el* 67.19% *ha aprobado un examen (el porcentaje ha sido redondeado al decimal exacto más próximo con dos cifras decimales). Determinar cuántos alumnos han aprobado y cuántos han suspendido.*

Solución: Sea A el número de aprobados y T el total de alumnos. Nos dicen que

$$\frac{100A}{T} \approx 67.19.$$

Más precisamente,

$$\left|\frac{6\,719}{10^2} - \frac{100A}{T}\right| \le \frac{1}{2\cdot 10^2},$$

pues todo número real dista a lo sumo $1/(2\cdot 10^2)$ de un decimal exacto con dos cifras decimales. Equivalentemente

$$\left|\frac{6\,719}{10^4} - \frac{A}{T}\right| \le \frac{1}{2\cdot 10^4} < \frac{1}{2T^2},$$

pues la última desigualdad equivale a $2T^2 < 2\cdot 10^4$, que se cumple, ya que $T < 100$. El teorema anterior (con la observación posterior) implica que A/T es un convergente de

$$\alpha = \frac{6\,719}{10^4} = [0,1,2,20,1,8,1,4,3]$$

El desarrollo en fracción continua se obtiene con el algoritmo de Euclides:

	6 719
0	10 000
1	6 719
2	3 281
20	157
1	141
8	16
1	13
4	3
3	1
	0

A su vez de aquí obtenemos los convergentes:

n	-1	0	1	2	3	4	5
a_n	$-$	0	1	2	20	1	8
p_n	1	0	1	2	41	43	385
q_n	0	1	1	3	61	64	573

Vemos que el mejor convergente A/T que cumple $T < 100$ es el dado por $A = 43$ y $T = 64$. Con estos valores, el porcentaje de aprobados es

$$\frac{100A}{T} = 67.1875\ldots \approx 67.19\%$$

y es la única solución, pues los convergentes anteriores nos darían

$$\frac{100 \cdot 41}{61} = 67.2131\ldots, \qquad \frac{100 \cdot 2}{3} = 66.6666\ldots$$

que no cumplen lo requerido. Concluimos que ha habido 43 aprobados y 21 suspensos. ∎

Ejercicio: En un pueblo de menos de mil habitantes hay un 44.8541% de hombres (donde el porcentaje está redondeado al decimal exacto más próximo con cuatro cifras decimales). Determinar cuántos hombres y cuántas mujeres hay en el pueblo.

10.2 Fracciones finalmente periódicas

Ejemplo En la sección precedente hemos calculado

$$\sqrt{2} = [1, 2, 2, 2, \ldots]$$

Puede decirse que nos hemos apoyado en un "truco", pero podríamos haber llegado a ella por un procedimiento completamente mecánico, sin más que ir aplicando el proceso determinado por

$$\alpha_0 = \alpha, \quad a_n = E[\alpha_n], \quad \alpha_{n+1} = \frac{1}{\alpha_n - \alpha}.$$

Lo resumimos en la tabla siguiente:

n	α_n	a_n
0	$\sqrt{2} \approx 1.41$	1
1	$1+\sqrt{2} \approx 2.41$	2
2	$1+\sqrt{2} \approx 2.41$	2

Para calcular la fila correspondiente a $n = 1$ hemos usado que

$$\frac{1}{\sqrt{2}-1} = \frac{\sqrt{2}+1}{(\sqrt{2}-1)(\sqrt{2}+1)} = 1 + \sqrt{2}$$

y, como $a_0 = 2$, para pasar a la fila siguiente tenemos que calcular

$$\frac{1}{1+\sqrt{2}+1} = \frac{1}{\sqrt{2}-1} = \frac{\sqrt{2}+1}{(\sqrt{2}-1)(\sqrt{2}+1)} = 1 + \sqrt{2}.$$

Puesto que $\alpha_2 = \alpha_1$, es claro que a partir de aquí todos los α_n serán este mismo número y todos los a_n serán iguales a 2. ∎

Ejemplo Si repetimos el proceso descrito en el ejemplo anterior para $\sqrt{23}$ el resultado es:

n	α_n	a_n
0	$\sqrt{23} \approx 4.79$	4
1	$\frac{4+\sqrt{23}}{7} \approx 1.25$	1
2	$\frac{3+\sqrt{23}}{2} \approx 3.89$	3
3	$\frac{3+\sqrt{23}}{7} \approx 1.11$	1
4	$4+\sqrt{23} \approx 8.79$	8
5	$\frac{4+\sqrt{23}}{7} \approx 1.25$	1

Como hemos llegado a que $\alpha_5 = \alpha_1$, resulta que $a_5 = a_1$, pero α_6 se calcula a partir de α_5 igual que α_2 se calcula a partir de α_1, luego necesariamente $\alpha_6 = \alpha_2$ y $a_6 = a_2$, y así sucesivamente, luego

$$\sqrt{23} = [4, 1, 3, 1, 8, 1, 3, 1, 8, 1, 3, 1, 8, \ldots].$$

En general, es claro que para que la fracción continua de un número irracional α sea finalmente periódica basta con que existan n y $h > 0$ tales que $\alpha_n = \alpha_{n+h}$. ∎

En la práctica usaremos la notación

$$\sqrt{2} = [1, \overline{2}], \qquad \sqrt{23} = [4, \overline{1, 3, 1, 8}]$$

para expresar que las cifras bajo la raya se repiten periódicamente. Notemos que la expresión no es única, pues, por ejemplo,

$$[4, \overline{1, 3, 1, 8}] = [4, \overline{1, 3, 1, 8, 1, 3, 1, 8}] = [4, 1, 3, \overline{1, 8, 1, 3}],$$

etc.

A las fracciones continuas de este tipo, que constan de un periodo repetido cíclicamente tal vez precedido por un anteperiodo, las llamaremos fracciones continuas finalmente periódicas.

En el cálculo de los desarrollos en fracción continua de $\sqrt{2}$ y $\sqrt{23}$ hemos tenido la "suerte" de que un α_n ha coincidido con un α_m precedente, lo que, según hemos señalado, se traduce en que la fracción continua resulta ser finalmente periódica y así acabamos determinando la sucesión por completo tras un número finito de pasos. El lector puede comprobar que la "suerte" le acompañará si intenta calcular los desarrollos en fracción continua de otras raíces cuadradas, como las que incluye la tabla 10.1. En todos los casos se encontrará con que un α_n se acaba repitiendo (concretamente, α_1). Si el lector observa la tabla con atención, encontrará algunos patrones adicionales.

Obviamente, esta "suerte" no es casual. Euler demostró que las fracciones continuas finalmente periódicas representan irracionales cuadráticos (es decir, raíces de polinomios de segundo grado con coeficientes enteros), y Lagrange demostró que el desarrollo en fracción continua de un irracional cuadrático es siempre finalmente periódico. Éstos son los resultados principales que demostraremos en esta sección.

Resumen 10.1: Desarrollos en fracción continua de raíces cuadradas

$\sqrt{2} = [1, \overline{2}]$	$\sqrt{11} = [3, \overline{3, 6}]$	$\sqrt{19} = [4, \overline{2, 1, 3, 1, 2, 8}]$
$\sqrt{3} = [1, \overline{1, 2}]$	$\sqrt{12} = [3, \overline{2, 6}]$	$\sqrt{20} = [4, \overline{2, 8}]$
$\sqrt{5} = [2, \overline{4}]$	$\sqrt{13} = [3, \overline{1, 1, 1, 1, 6}]$	$\sqrt{21} = [4, \overline{1, 1, 2, 1, 1, 8}]$
$\sqrt{6} = [2, \overline{2, 4}]$	$\sqrt{14} = [3, \overline{1, 2, 1, 6}]$	$\sqrt{22} = [4, \overline{1, 2, 4, 2, 1, 8}]$
$\sqrt{7} = [2, \overline{1, 1, 1, 4}]$	$\sqrt{15} = [3, \overline{1, 6}]$	$\sqrt{23} = [4, \overline{1, 3, 1, 8}]$
$\sqrt{8} = [2, \overline{1, 4}]$	$\sqrt{17} = [4, \overline{8}]$	$\sqrt{24} = [4, \overline{1, 8}]$
$\sqrt{10} = [3, \overline{6}]$	$\sqrt{18} = [4, \overline{4, 8}]$	$\sqrt{26} = [5, \overline{10}]$

Ejercicio: Probar que $\sqrt{c^2+1} = [c, \overline{2c}]$, $\sqrt{c^2+2} = [c, \overline{c, 2c}]$.

Empezamos probando el resultado de Euler, que se deduce fácilmente del teorema siguiente:

Teorema 10.10 *Sea* $\alpha = [a_0, a_1, a_2, \dots]$ *y sea* $\beta = [a_{n+1}, a_{n+2}, a_{n+3}, \dots]$, *para* $n \geq 1$. *Entonces se cumple que*

$$\alpha = \frac{\beta p_n + p_{n-1}}{\beta q_n + q_{n-1}}.$$

Demostración: La prueba consiste simplemente en observar que en la demostración del teorema 10.2 no se ha usado que los coeficientes a_n sean enteros salvo para probar que $(p_n, q_n) = 1$. Por lo tanto podemos aplicarlo a $\alpha = [a_0, \dots, a_n, \beta]$ y concluir que, aunque ahora p_{n+1} y q_{n+1} no sean números racionales,

$$\alpha = \frac{p_{n+1}}{q_{n+1}} = \frac{\beta p_n + p_{n-1}}{\beta q_n + q_{n-1}}$$ ∎

Por ejemplo, si $\alpha = [1, 2, 3, \overline{4, 5}]$ y llamamos $\beta = [\overline{4, 5}]$, tenemos la relación $\alpha = [1, 2, 3, \beta]$, luego

n	-1	0	1	2	3
a_n	$-$	1	2	3	β
p_n	1	1	3	10	$10\beta + 3$
q_n	0	1	2	7	$7\beta + 2$

por lo que

$$\alpha = \frac{10\beta + 3}{7\beta + 2}.$$

Esto nos reduce el cálculo de α al caso de una fracción continua periódica, pero ahora podemos aplicar el teorema a la relación $\beta = [4, 5, \beta]$. Para ello calculamos

n	-1	0	1	2
a_n	$-$	4	5	β
p_n	1	4	21	$21\beta + 4$
q_n	0	1	5	$5\beta + 1$

con lo que

$$\beta = \frac{21\beta + 4}{5\beta + 1}$$

luego

$$5\beta^2 + \beta = 21\beta + 4$$

o también

$$5\beta^2 - 20\beta - 4 = 0.$$

Teniendo en cuenta que $\beta > 0$, concluimos que

$$\beta = \frac{10 + 2\sqrt{30}}{5},$$

luego

$$\alpha = \frac{10\beta + 3}{7\beta + 2} = \frac{80 - \sqrt{30}}{52}.$$

Este procedimiento puede aplicarse a cualquier fracción continua finalmente periódica, luego hemos demostrado que todas ellas determinan irracionales cuadráticos.

Ejemplo Vamos a calcular $\alpha = [1, 1, 1, \ldots]$.

Para ello usamos que $\alpha = [1, \alpha]$, luego tenemos que

n	-1	0	1
a_n	$-$	1	α
p_n	1	1	$\alpha + 1$
q_n	0	1	α

Por lo tanto,

$$\alpha = \frac{\alpha + 1}{\alpha},$$

y así $\alpha^2 - \alpha - 1 = 0$, de donde $\alpha = \frac{1+\sqrt{5}}{2}$. Así pues:

$$\frac{1 + \sqrt{5}}{2} = 1 + \frac{1}{1 + \frac{1}{1 + \cdots}}$$

El número $\Phi = \frac{1+\sqrt{5}}{2}$ se conoce como *número áureo*, que ya había llamado la atención de matemáticos y artistas en la antigüedad.

Ejercicio: Encontrar la proporción que deben tener los lados de un rectángulo (el mayor sobre el menor) para que el rectángulo lateral que muestra la figura guarde la misma proporción.

En geometría aparece en el estudio del pentágono, y nosotros lo hemos encontrado como la unidad fundamental del cuerpo $\mathbb{Q}(\sqrt{5})$. Otra propiedad se obtiene al calcular sus convergentes:

n	-1	0	1	2	3	4
a_n	$-$	1	1	1	1	1
p_n	1	1	2	3	5	8
q_n	0	1	1	2	3	5

Es claro que la sucesión q_n está determinada por las relaciones

$$q_0 = q_1 = 1, \qquad q_n = q_{n-1} + q_{n-2},$$

mientras que $p_n = q_{n+1}$. La sucesión

$$1, \quad 1, \quad 2, \quad 3, \quad 5, \quad 8, \quad 13, \quad 21, \quad \ldots$$

en la que cada término es la suma de los dos precedentes se conoce como *sucesión de Fibonacci.* Ahora sabemos que los cocientes de dos términos consecutivos de la sucesión de Fibonacci son los convergentes del número áureo, luego en particular convergen hacia él. ∎

Consideremos ahora el caso de las fracciones continuas periódicas, es decir, las que no tienen anteperiodo. Es fácil dar condiciones necesarias para que un irracional cuadrático pueda tener una fracción continua periódica:

Teorema 10.11 *Si un número α tiene fracción continua periódica, entonces α es un irracional cuadrático que cumple $\alpha > 1$ y $-1 < \bar{\alpha} < 0$.*

Demostración: Si $\alpha = [\overline{a_0, \ldots, a_n}]$, entonces α_n tiene la misma fracción continua que α, luego por la unicidad $\alpha = \alpha_n > 1$. El teorema 10.10 nos da que

$$\alpha = \frac{p_n\alpha + p_{n-1}}{q_n\alpha + q_{n-1}},$$

luego $(q_n\alpha + q_{n-1})\alpha = p_n\alpha + p_{n-1}$. Esto hace que α sea raíz del polinomio

$$f(x) = q_n x^2 + (q_{n-1} - p_n)x - p_{n-1} = q_n(x - \alpha)(x - \bar{\alpha}).$$

Por lo tanto

$$q_n\alpha\bar{\alpha} = f(0) = -p_{n-1} < 0,$$

$$q_n(-1 - \alpha)(-1 - \bar{\alpha}) = f(-1) = q_n - q_{n-1} + p_n - p_{n-1} > 0,$$

La primera desigualdad nos da que $\bar{\alpha} < 0$ y la segunda que $-1 - \bar{\alpha} < 0$. ∎

Lo relevante es que el recíproco del teorema anterior también es cierto. Para probarlo conviene definir un irracional cuadrático *reducido* como un irracional cuadrático α que cumpla las condiciones del teorema anterior, es decir, que $\alpha > 1$ y $-1 < \bar{\alpha} < 0$.

Por ser un irracional cuadrático, será raíz de un polinomio con coeficientes enteros:

$$ax^2 + bx + c.$$

donde podemos suponer que $a > 0$. Las raíces son

$$\frac{-b \pm \sqrt{\Delta}}{2a}, \qquad \Delta = b^2 - 4ac > 0,$$

pero, como $-1 < \bar{\alpha} < 0 < 1 < \alpha$, necesariamente

$$\alpha = \frac{B + \sqrt{\Delta}}{A}, \qquad B = -b, \quad A = 2a > 0.$$

Más aún, como $\alpha > 1$ y $\bar{\alpha} > -1$, se cumple que $\alpha + \bar{\alpha} > 0$, pero esto equivale a que $2B/A > 0$, luego también $B > 0$.

Más aún, como $\bar{\alpha} < 0$, se cumple que $B - \sqrt{\Delta} < 0$, luego $B < \sqrt{D}$ y, como $\alpha > 1$, también tenemos que $B + \sqrt{\Delta} > A$, luego $A < 2\sqrt{\Delta}$. En resumen:

> *Todo irracional cuadrático reducido α puede expresarse en la forma*
>
> $$\alpha = \frac{B + \sqrt{\Delta}}{A},$$
>
> *donde A, B, Δ son enteros, pero entonces necesariamente $\Delta > 0$ y $0 < A < 2\sqrt{\Delta}$, $0 < B < \sqrt{\Delta}$. En particular, hay un número finito de irracionales cuadráticos reducidos expresables en términos de un mismo discriminante Δ.*

Con esto ya podemos probar:

Teorema 10.12 *El desarrollo en fracción continua de un número real α es periódico si y sólo si α es un irracional cuadrático reducido.*

Demostración: Ya hemos probado una implicación. Supongamos que α es un irracional cuadrático reducido, al que podemos aplicar, por tanto, toda la discusión precedente, y vamos a estudiar su desarrollo en fracción continua. Para ello empezamos calculando $a_0 = E[\alpha]$ y $\alpha_1 = 1/(\alpha - a_0)$, de modo que

$$\alpha = a_0 + \frac{1}{\alpha_1}.$$

Vamos a ver que α_1 también es un irracional cuadrático reducido. Claramente es irracional (o α sería racional) y podemos razonar que es cuadrático porque pertenece al mismo cuerpo cuadrático al cual pertenece α, pero, de todos modos, enseguida calcularemos explícitamente un polinomio de segundo grado del cual es raíz.

Como $1/\alpha_1$ es la parte fraccionaria de α, se cumple que $0 < 1/\alpha_1 < 1$, luego $\alpha_1 > 1$. Por otra parte, como $\alpha > 1$, se cumple que $a_0 = E(\alpha) \geq 1$, luego

$$\frac{1}{\bar{\alpha}_1} = \bar{\alpha} - a_0 < 0 - 1 = -1,$$

luego $\bar{\alpha}_1 < 0$ y a su vez esto y la desigualdad precedente implican que $-1 < \bar{\alpha}$.

Esto prueba que α_1 es reducido, y sustituyendo en la ecuación de α vemos que

$$a\left(a_0 + \frac{1}{\alpha_1}\right)^2 + b\left(a_0 + \frac{1}{\alpha_1}\right) + c = 0.$$

Desarrollando queda

$$(aa_0^2 + ba_0 + c)\alpha_1^2 + (2aa_0 + b)\alpha_1 + a = 0.$$

Notemos que $aa_0^2 + ba_0 + c \neq 0$, pues las raíces del polinomio $ax^2 + bx + c$ son α y $\bar{\alpha}$, que no son enteras, luego no son a_0. Es fácil ver que el discriminante de este polinomio es también Δ, con lo que

$$\alpha_1 = \frac{B_1 + \sqrt{\Delta}}{A_1}$$

en las mismas condiciones que α, y exactamente con el mismo discriminante Δ.

Por el mismo argumento, resulta que lo mismo le sucede a α_2 y, en general, a todos los α_n. Todos son irracionales cuadráticos reducidos expresables en términos del mismo discriminante Δ, pero hemos visto que sólo hay un número finito de tales irracionales, así que, tras un número finito de pasos, tiene que darse una coincidencia $\alpha_k = \alpha_l$, con $k < l$. Esto nos asegura que el desarrollo en fracción continua es finalmente periódico, pero vamos a ver que es periódico.

Si $k > 0$, tenemos que

$$\alpha_{k-1} = a_{k-1} + \frac{1}{\alpha_k}, \qquad \alpha_{l-1} = a_{l-1} + \frac{1}{\alpha_l},$$

luego

$$\bar{\alpha}_{k-1} = a_{k-1} + \frac{1}{\bar{\alpha}_k}, \qquad \bar{\alpha}_{l-1} = a_{l-1} + \frac{1}{\bar{\alpha}_l},$$

luego

$$a_{k-1} - \bar{\alpha}_{k-1} = -\frac{1}{\bar{\alpha}_k} = -\frac{1}{\bar{\alpha}_l} = a_{l-1} - \bar{\alpha}_{l-1}.$$

Pero a_{k-1} y a_{l-1} son números naturales y $-\bar{\alpha}_{k-1}$ y $-\bar{\alpha}_{l-1}$ son números estrictamente comprendidos entre 0 y 1. Por la unicidad de la parte entera, concluimos que $a_{k-1} = a_{l-1}$ y que $\bar{\alpha}_{k-1} = \bar{\alpha}_{l-1}$, luego $\alpha_{k-1} = \alpha_{l-1}$. Repitiendo este proceso llegamos a que $\alpha_0 = \alpha_{l-k}$ y así el desarrollo de α es periódico. ∎

Ejemplo Si n es un número natural que no sea un cuadrado perfecto, entonces $\alpha = \sqrt{n}$ no es reducido, ya que $\bar{\alpha} = -\sqrt{n} < -1$, luego el desarrollo en fracción continua de $\sqrt{n}$ no puede ser periódico. Sin embargo, si $a_0 = E(\sqrt{n}) \geq 1$, entonces $\alpha^* = a_0 + \sqrt{n}$ sí que es reducido, pues ciertamente $\alpha^* > 1$ y, como

$$a_0 < \sqrt{n} < a_0 + 1,$$

también

$$-1 < \bar{\alpha}^* = a_0 - \sqrt{n} < 0.$$

Además $E[\alpha^*] = 2a_0$, luego, por el teorema anterior,

$$a_0 + \sqrt{n} = [\overline{2a_0, a_1, \ldots, a_n}] = [2a_0, \overline{a_1, \ldots a_n, 2a_0}],$$

y es claro entonces que

$$\sqrt{n} = [a_0, \overline{a_1, \ldots, a_n, 2a_0}].$$

Esto prueba que, tal y como se veía en la tabla 10.1, las raíces cuadradas tienen desarrollos finalmente periódicos con anteperiodo de longitud 1 igual a la mitad del último coeficiente del periodo. ∎

Finalmente podemos probar:

Teorema 10.13 *Las fracciones continuas finalmente periódicas se corresponden con los irracionales cuadráticos.*

DEMOSTRACIÓN: Ya hemos probado una implicación. Supongamos que α es un irracional cuadrático y veamos que su fracción continua es finalmente periódica. El teorema 10.10 nos da que, para todo $n \geq 1$, se cumple la relación

$$\alpha = \frac{\alpha_{n+1}p_n + p_{n-1}}{\alpha_{n+1}q_n + q_{n-1}},$$

de donde

$$\bar{\alpha} = \frac{\bar{\alpha}_{n+1}p_n + p_{n-1}}{\bar{\alpha}_{n+1}q_n + q_{n-1}}.$$

Despejando:

$$\bar{\alpha}_{n+1} = -\frac{\bar{\alpha}q_{n-1} - p_{n-1}}{\bar{\alpha}q_n - p_n} = -\frac{q_{n-1}}{q_n}\,\frac{\bar{\alpha} - r_{n-1}}{\bar{\alpha} - r_n} = -\frac{q_{n-1}}{q_n}\,\frac{r_{n-1} - \bar{\alpha}}{r_n - \bar{\alpha}},$$

donde $r_r = p_n/q_n$ es el convergente n-simo.

Si $\bar{\alpha} > \alpha$ tomamos n par, y en caso contrario lo tomamos impar. Vamos a considerar el primer caso, pero el segundo nos lleva análogamente a la misma conclusión. El teorema 10.4 nos da que $r_n < \alpha < r_{n-1}$, luego

$$\bar{\alpha} - r_{n-1} = \bar{\alpha} - \alpha - (r_{n-1} - \alpha)$$

y, tomando n suficientemente grande, podemos hacer que $r_{n-1} - \alpha < \bar{\alpha} - \alpha$, de modo que

$$0 < \bar{\alpha} - r_{n-1} < \bar{\alpha} - \alpha,$$

y, por otra parte, $\bar{\alpha} - r_n > \bar{\alpha} - \alpha > 0$, luego

$$0 < \frac{\bar{\alpha} - r_{n-1}}{\bar{\alpha} - r_n} < \frac{\bar{\alpha} - \alpha}{\bar{\alpha} - \alpha} = 1$$

y, como $0 < q_{n-1}/q_n < 1$, concluimos que

$$-1 < \bar{\alpha}_{n+1} < 0.$$

Por otro lado, la condición $\alpha_{n+1} > 1$ se da siempre, luego resulta que α_{n+1} es un irracional cuadrático reducido para todo n (de la paridad adecuada) suficientemente grande. Esto implica que la fracción continua de α_{n+1} es periódica, luego la de α es finalmente periódica. ∎

La tabla 10.1 muestra una propiedad adicional de los desarrollos de raíces cuadradas en fracciones continuas que no es difícil de probar: los periodos, salvo por su último coeficiente (que ya sabemos que es el doble del anteperiodo) son simétricos (capicúas). Esto es consecuencia del teorema siguiente:

Teorema 10.14 *Sea $\alpha = [\overline{a_0, \ldots, a_n}]$ un irracional cuadrático reducido y sea $\beta = [\overline{a_n, \ldots, a_0}]$. Entonces $\bar{\alpha} = -1/\beta$.*

DEMOSTRACIÓN: Llamemos p_k/q_k a los convergentes de α y p_k^*/q_k^* a los convergentes de β. Tenemos las relaciones

$$p_k = a_k p_{k-1} + p_{k-2}, \qquad q_k = a_k q_{k-1} + q_{k-2},$$

luego

$$\frac{p_k}{p_{k-1}} = a_k + \frac{1}{p_{k-1}/p_{k-2}}, \qquad \frac{q_k}{q_{k-1}} = a_k + \frac{1}{q_{k-1}/q_{k-2}}.$$

En particular:

$$\frac{p_1}{p_0} = a_1 + \frac{1}{p_0/p_{-1}} = a_1 + \frac{1}{a_0} = [a_1, a_0],$$

luego

$$\frac{p_2}{p_1} = a_2 + \frac{1}{p_1/p_0} = a_2 + \frac{1}{[a_1, a_0]} = [a_2, a_1, a_0],$$

y así llegamos a que

$$\frac{p_n}{p_{n-1}} = [a_n, \ldots a_0] = \frac{p_n^*}{q_n^*}.$$

Igualmente,

$$\frac{q_1}{q_0} = a_1 = [a_1],$$

luego

$$\frac{q_2}{q_1} = a_2 + \frac{1}{q_1/q_0} = a_2 + \frac{1}{a_1} = [a_2, a_1],$$

y así llegamos a que

$$\frac{q_n}{q_{n-1}} = [a_n, \ldots, a_1] = \frac{p_{n-1}^*}{q_{n-1}^*}.$$

Como las fracciones son irreducibles, concluimos que

$$p_n^* = p_n, \quad p_{n-1}^* = q_n, \quad q_n^* = p_{n-1}, \quad q_{n-1}^* = q_{n-1}.$$

Como $\alpha = \alpha_{n+1}$ y $\beta = \beta_{n+1}$, el teorema 10.10 nos da que

$$\alpha = \frac{\alpha p_n + p_{n-1}}{\alpha q_n + q_{n-1}}, \quad \beta = \frac{\beta p_n^* + p_{n-1}^*}{\beta q_n^* + q_{n-1}^*} = \frac{\beta p_n + q_n}{\beta p_{n-1} + q_{n-1}}.$$

Operando vemos que α y $-1/\beta$ satisfacen la misma ecuación cuadrática, pero uno es positivo y el otro negativo, luego no son iguales, luego necesariamente $-1/\beta = \bar{\alpha}$. ∎

Teorema 10.15 *Si n es un número natural que no es un cuadrado perfecto, entonces*

$$\sqrt{n} = [a_0, \overline{a_1, \ldots, a_n, 2a_0}],$$

donde la sucesión $a_1, \ldots, a_n$ es capicúa.

DEMOSTRACIÓN: En el ejemplo tras el teorema 10.12 hemos visto que $a_0 + \sqrt{n}$ es un irracional cuadrático reducido, de modo que

$$a_0 + \sqrt{n} = [\overline{2a_0, a_1, \ldots, a_n}], \qquad \sqrt{n} = [2a_0, \overline{a_1, \ldots, a_n, 2a_0}],$$

luego el teorema anterior nos da que

$$\frac{1}{\sqrt{n} - a_0} = [\overline{a_n, \ldots, a_1, 2a_0}].$$

Pero, si $\alpha = \sqrt{n}$, entonces el miembro izquierdo de la última igualdad es α_1, luego

$$\frac{1}{\sqrt{n} - a_0} = [\overline{a_1, \ldots, a_n, 2a_0}].$$

Comparando los dos desarrollos concluimos que $a_n, \ldots, a_1$ es la misma sucesión que $a_1, \ldots, a_n$, luego es capicúa. ∎

10.3 La ecuación de Pell

Ya estamos en condiciones de resolver cómodamente la ecuación de Pell:

$$x^2 - Dy^2 = 1,$$

donde D es un número natural que no sea un cuadrado perfecto.[3] Descompongamos $D = m^2 d$, donde d es libre de cuadrados. Así $\sqrt{D} = m\sqrt{d}$.

Definamos $\mathbb{Z}[m\sqrt{d}]$ como el conjunto de todos los números $a + bm\sqrt{d}$, donde a y b son enteros racionales. Es claro que se trata de un subanillo del anillo de enteros $\mathbb{Z}[\omega]$ del cuerpo $\mathbb{Q}(\sqrt{d})$ (esto es poco más que afirmar que la suma y el producto de elementos de $\mathbb{Z}[m\sqrt{d}]$ está en $\mathbb{Z}[m\sqrt{d}]$, como se comprueba sin dificultad).

Si (x, y) es una solución entera de la ecuación $x^2 - Dy^2 = \pm 1$, entonces

$$(x + ym\sqrt{d})(x - ym\sqrt{d}) = x^2 - y^2 m^2 d = \pm 1,$$

luego $\epsilon = x + ym\sqrt{d}$ es una unidad de $\mathbb{Z}[m\sqrt{d}]$ (y en particular es una unidad de $\mathbb{Z}[\omega]$). Recíprocamente, cada unidad de $\mathbb{Z}[m\sqrt{d}]$ determina una solución de la ecuación $x^2 - Dy^2 = \pm 1$.

[3] Si $D = c^2$, la ecuación es $(x + cy)(x - cy) = 1$, que equivale a $x + cy = \pm 1$, $x - cy = \pm 1$, de donde, sumando las ecuaciones, $2x = \pm 2$, luego $x = \pm 1$, luego $y = 0$, y así las únicas soluciones son $(\pm 1, 0)$.

Sea η la unidad fundamental de $\mathbb{Z}[\omega]$. Vamos a probar que existe un número natural $k > 0$ tal que las unidades de $\mathbb{Z}[\omega]$ son las potencias de $\epsilon = \eta^k$.

El anillo de clases de restos $\mathbb{Z}[\omega]_{2m}$ es finito, pues, dado un entero arbitrario $a + b\omega$, podemos expresar $a = 2mc_1 + r_1$, $b = 2mc_2 + r_2$, con $0 \le r_1, r_2 < 2m$, y entonces

$$a + b\omega = 2m(c_1 + c_2\omega) + r_1 + r_2\omega \equiv r_1 + r_2\omega \pmod{2m},$$

luego $\mathbb{Z}[\omega]_{2m}$ tiene a lo sumo $4m^2$ elementos.

Ejercicio: Probar que $\mathbb{Z}[\omega]_{2m}$ tiene exactamente $4m^2$ elementos.

Sea W_{2m} el grupo de unidades de $\mathbb{Z}[\omega]_{2m}$, que también es finito. Uno de sus elementos es la clase $\bar\eta$, y tendrá orden finito en W_{2m}, lo que se traduce en que existe un número natural $k > 0$ tal que $\bar\eta^k = 1$, es decir, que $\eta^k \equiv 1 \pmod{2m}$, luego

$$\eta^k - 1 = m(2a + b2\omega) = u + vm\sqrt{d}.$$

(el 2 sólo es necesario en el caso en que $d \equiv 1 \pmod 4$, pues así $2\omega = 1 + \sqrt{d}$). Concluimos que η^k está en $\mathbb{Z}[m\sqrt{d}]$. Tomemos el mínimo $k > 0$ tal que $\epsilon = \eta^k$ está en $\mathbb{Z}[m\sqrt{d}]$ y vamos a ver que cumple lo requerido.

Ciertamente, las potencias de ϵ son unidades de $\mathbb{Z}[m\sqrt{d}]$. Recíprocamente, si ϵ' es una unidad arbitraria de $\mathbb{Z}[m\sqrt{d}]$, también lo es de $\mathbb{Z}[\omega]$, luego $\epsilon' = \eta^s$, para cierto entero s, que podemos expresar como $s = kc + r$, con $0 \le r < k$. Entonces $\epsilon' = \eta^{kc}\eta^r = \epsilon^c\eta^r$, luego $\eta^r = \epsilon'\epsilon^{-c}$ está en $\mathbb{Z}[m\sqrt{d}]$, luego por la minimalidad de k tiene que ser $r = 0$, y así $\epsilon' = \epsilon^c$.

Con esto hemos demostrado el teorema siguiente:

Teorema 10.16 *Si D es un número natural no cuadrado perfecto, las unidades del anillo $\mathbb{Z}[\sqrt{D}\,]$ son las potencias de una unidad fundamental $\epsilon > 1$, de modo que, si $\mathrm{N}(\epsilon) = 1$, sus potencias $\epsilon^n = x_n + y_n\sqrt{D}$, con n entero, determinan todas las soluciones (x_n, y_n) de la ecuación de Pell*

$$x^2 - Dy^2 = 1,$$

mientras que la ecución $x^2 - Dy^2 = -1$ no tiene soluciones enteras. En cambio, si $\mathrm{N}(\epsilon) = -1$, las potencias pares de ϵ determinan las soluciones de la ecuación de Pell y las potencias impares las de la ecuación $x^2 - Dy^2 = -1$.

Ahora sólo nos falta dar un procedimiento para encontrar la unidad fundamental de un anillo $\mathbb{Z}[\sqrt{D}\,]$. Si $\epsilon = x + y\sqrt{D} > 1$ es una unidad de $\mathbb{Z}[\sqrt{D}\,]$, entonces $\mathrm{N}(\epsilon) = \epsilon\bar\epsilon = \pm 1$, luego $\bar\epsilon = \pm 1/\epsilon$, de donde $\epsilon - \bar\epsilon > 0$, pero esto es $2y\sqrt{D} > 0$, luego $y \ge 1$. Por otra parte, $|x - y\sqrt{D}| = |\bar\epsilon| < 1$, luego $x \ge 1$. La relación

$$(x + y\sqrt{D})(x - y\sqrt{D}) = x^2 - Dy^2 = \pm 1$$

implica, por una parte, que $(x, y) = 1$ y, por otra

$$x - y\sqrt{D} = \pm\frac{1}{x + y\sqrt{D}},$$

luego

$$\frac{x}{y} - \sqrt{D} = \pm\frac{1}{y(x + y\sqrt{D})}$$

y así

$$\left|\frac{x}{y} - \sqrt{D}\right| = \frac{1}{y(x + y\sqrt{D})} \le \frac{1}{y(y\sqrt{D-1} + y\sqrt{D})} = \frac{1}{y^2(\sqrt{D-1} + \sqrt{D})} < \frac{1}{2y^2},$$

donde hemos usado que

$$x^2 = Dy^2 \pm 1 \ge Dy^2 - 1 \ge Dy^2 - y^2 = y^2(D-1),$$

luego $x \ge y\sqrt{D-1}$. Como $(x, y) = 1$, el teorema 10.9 implica que x/y es un convergente de $\sqrt{D}$.

Ahora, si ϵ es la unidad fundamental y $\epsilon^n = x_n + y_n\sqrt{D}$, es fácil ver que las sucesiones x_n e y_n son estrictamente crecientes (usando que $x_n, y_n > 0$), y también lo son las sucesiones p_n, q_n que determinan los convergentes, luego concluimos que la unidad fundamental $x_1 + y_1\sqrt{D}$ se corresponde con el primer convergente de $\sqrt{D}$ que cumple $p_n^2 - Dq_n^2 = \pm 1$. Recogemos esto en el teorema siguiente:

Teorema 10.17 *Si D es un número natural no cuadrado perfecto, entonces la unidad fundamental de $\mathbb{Z}[\sqrt{D}]$ es $\epsilon = x + y\sqrt{D}$, donde x/y es el menor convergente de $\sqrt{D}$ que cumple $x^2 - Dy^2 = \pm 1$.*

Ejemplo Vamos a resolver las ecuaciones

$$x^2 - 13y^2 = 1, \qquad x^2 - 13y^2 = -1.$$

Para ello calculamos el desarrollo en fracción continua

$$\sqrt{13} = [3, \overline{1, 1, 1, 1, 6}]$$

y a partir de él calculamos sus convergentes:

n	-1	0	1	2	3	4
a_n	$-$	3	1	1	1	1
p_n	1	3	4	7	11	18
q_n	0	1	1	2	3	5
$p_n^2 - 13q_n^2$	$-$	-4	3	4	4	-1

Concluimos que la unidad fundamental de $\mathbb{Z}[\sqrt{13}]$ es $\epsilon = 18 + 5\sqrt{13}$, que tiene norma negativa, luego la menor solución no trivial de la ecuación de Pell es la dada por $\epsilon^2 = 649 + 180\sqrt{13}$. Las demás se obtienen a partir de las unidades

$$(649 + 180\sqrt{13})^n.$$

Por otro lado, las soluciones de la ecuación $x^2 - 13y^2 = -1$ son las de la forma

$$(18 + 5\sqrt{13})(649 + 180\sqrt{13})^n.$$

∎

El acertijo del rey Harold Vamos a calcular ahora el (menor posible) presunto número de soldados del rey Harold según el acertijo de Dudeney que hemos planteado en la introducción de este capítulo. Éste está determinado por la ecuación de Pell

$$x^2 - 61y^2 = 1.$$

Necesitamos el desarrollo

$$\sqrt{61} = [7, \overline{1, 4, 3, 1, 2, 2, 1, 3, 4, 1, 14}].$$

A partir de él calculamos los convergentes:

n	-1	0	1	2	3	4	5	6	7	8	9	10
a_n	$-$	7	1	4	3	1	2	2	1	3	4	14
p_n	1	7	8	39	125	164	453	1 070	1 523	5 639	24 079	29 718
q_n	0	1	1	5	16	21	58	137	195	722	3 083	3 805
$p_n^2 - 13q_n^2$	$-$	-12	3	-4	9	-5	5	-9	4	-3	12	-1

Así pues, la unidad fundamental de $\mathbb{Z}\big[\sqrt{61}\,\big]$ es $\epsilon = 29\,718 + 3\,805\sqrt{61}$, pero tiene norma negativa, luego la solución fundamental de la ecuación de Pell es la determinada por su cuadrado:

$$\epsilon^2 = 1\,766\,319\,049 + 226\,153\,980\sqrt{61}.$$

Por lo tanto, el ejército de Harold constaba supuestamente de 61 divisiones de

$$(226\,153\,980)^2 = 51\,145\,622\,669\,840\,400$$

soldados cada una, que junto con Harold sumaban

$$(1\,766\,319\,049)^2 = 3\,119\,882\,982\,860\,264\,401 \text{ soldados.}$$

Hay que advertir que el valor $D = 61$ está elegido con intención, pues muchas otras ecuaciones de Pell con parámetros similares admiten soluciones mucho menores. ∎

Ejercicio: Hallar las soluciones enteras de $x^2 - 60y^2 = 1$ y $x^2 - 62y^2 = 1$.

Ejercicio: Resolver la ecuación de Pell propuesta por Fermat: $x^2 - 109y^2 = 1$. El desarrollo en fracción continua es

$$\sqrt{109} = [10, \overline{2, 3, 1, 2, 4, 1, 6, 6, 1, 4, 2, 1, 3, 2, 20}].$$

Hay un caso en el que la solución fundamental de la ecuación de Pell se obtiene inmediatamente:

Teorema 10.18 *Si $D = c^2 - 1$, con $c \geq 2$, a solución fundamental de la ecuación de Pell $x^2 - Dy^2 = 1$ es $(c, 1)$.*

Demostración: Ciertamente es una solución, y la solución fundamental (a, b) debe cumplir $1 \leq b \leq 1$, luego $b = 1$, lo que fuerza que $a = c$. ∎

10.4 Unidades de órdenes cuadráticos

Los resultados de la sección anterior nos permiten calcular la unidad fundamental del anillo de enteros de cualquier cuerpo cuadrático real $\mathbb{Q}(\sqrt{d}\,)$ con $d \not\equiv 1 \pmod 4$. Sin embargo, refinando ligeramente el argumento podemos extenderlo al caso $d \equiv 1 \pmod 4$.

Un poco más en general, si $\mathbb{Q}(\sqrt{d}\,)$ es un cuerpo cuadrático y $\mathbb{Z}[\omega]$ es su anillo de enteros, definimos el *orden* de $\mathbb{Q}(\sqrt{d}\,)$ de índice m como el anillo $\mathcal{O}_m$ formado por los números de la forma $a+bm\omega$, donde a y b son enteros racionales.

Ejercicio: Comprobar que el conjugado de un elemento de $\mathcal{O}_m$ está en $\mathcal{O}_m$.

Notemos que $\mathcal{O}_1$ es el anillo de enteros de $\mathbb{Q}(\sqrt{d}\,)$.

Si $d \not\equiv 1 \pmod 4$, estos órdenes no son sino los anillos $\mathcal{O}_m = \mathbb{Z}[m\sqrt{d}]$ que hemos considerado en la sección precedente. En cambio, si $d \equiv 1 \pmod 4$ entonces $\mathbb{Z}[m\sqrt{d}] = \mathcal{O}_{2m}$. En efecto, los elementos de $\mathcal{O}_{2m}$ son los de la forma

$$a + b2m\omega = a + bm(1 + \sqrt{d}) = a + bm + bm\sqrt{d},$$

y es claro que los elementos de esta forma coinciden con los de la forma $a+bm\sqrt{d}$, es decir, con los de $\mathbb{Z}[m\sqrt{d}]$.

Aunque aquí estamos interesados en los órdenes de los cuerpos cuadráticos reales, observemos que todo lo dicho vale también si $d < 0$.

Ejercicio: Probar que si $d < 0$ y $m > 1$, las unidades del orden cuadrático $\mathcal{O}_m$ son únicamente ± 1.

A partir de aquí suponemos $d > 0$ y observamos que las unidades de $\mathcal{O}_m$ son de la forma ϵ^n, donde n recorre los números enteros, para cierta unidad fundamental $\epsilon > 0$. En efecto, en la sección anterior hemos demostrado que si η es la unidad fundamental de $\mathbb{Q}(\sqrt{d}\,)$, existe un exponente $k > 0$ tal que η^k está en $\mathcal{O}_{2m} \subset \mathcal{O}_m$, luego podemos considerar el menor exponente $k > 0$ tal que $\epsilon = \eta^k$ está en $\mathcal{O}_m$. Exactamente el mismo razonamiento empleado allí prueba que las unidades de $\mathcal{O}_m$ son las potencias de ϵ.

Teorema 10.19 *Sea $\mathcal{O}_m$ un orden de un cuerpo cuadrático real $\mathbb{Q}(\sqrt{d}\,)$ (con $m \geq 2$ si $d = 5$). Entonces, la unidad fundamental de $\mathcal{O}_m$ es $\epsilon = x + ym\omega$, donde x/y es el menor convergente de $-m\bar{\omega}$ que cumple $\mathrm{N}(x + ym\omega) = \pm 1$.*

Demostración: Si $d \not\equiv 1 \pmod 4$, tenemos que $-m\bar{\omega} = m\sqrt{d}$, por lo que el resultado está demostrado en 10.17. Por lo tanto, podemos restringirnos al caso en que $d \equiv 1 \pmod 4$.

Sea $\epsilon = x + ym\omega > 1$ una unidad arbitraria de $\mathcal{O}_m$. Por el teorema 9.5 sabemos que $x, y \geq 1$ (incluso si $d = 5$, pues entonces $m \geq 2$ y $\epsilon \neq \omega$). Como $\mathrm{N}(\epsilon) = (x + ym\omega)(x + ym\bar{\omega}) = \pm 1$, necesariamente $(x, y) = 1$ y además

$$\left|\frac{x}{y} - (-m\bar{\omega})\right| = \left|\frac{x}{y} + m\bar{\omega}\right| = \frac{1}{y(x + ym\omega)} \leq \frac{1}{y^2 m\omega} < \frac{1}{2y^2},$$

pues

$$m\omega = m\frac{1+\sqrt{d}}{2} > 2,$$

bien porque $d \geq 13$, bien porque $d = 5$ y $m \geq 2$. Como $(x, y) = 1$, el teorema 10.9 implica que x/y es un convergente de $-m\bar{\omega}$.

Si ϵ es la unidad fundamental de $\mathcal{O}_m$ y $\epsilon^n = x_n + y_n m\omega$, las sucesiones x_n y my_n son subsucesiones de las consideradas en el teorema 9.8, por lo que podemos concluir que son estrictamente crecientes (incluso si $d = 5$, por el ejemplo precedente al teorema citado, ya que las excepciones a la monotonía no están en $\mathcal{O}_m$ si $m \geq 2$). Por lo tanto, la unidad fundamental está asociada al primer convergente p_n/q_n de $-m\bar{\omega}$ que cumple $\mathrm{N}(p_n + q_n m\omega) = 1$. ∎

Ejemplo Vamos a calcular la unidad fundamental del anillo de enteros de $\mathbb{Q}\big(\sqrt{61}\,\big)$. Para ello necesitamos el desarrollo

$$-\frac{1-\sqrt{61}}{2} = [3, \overline{2, 2, 7}],$$

con el que podemos calcular los convergentes:

n	-1	0	1	2
a_n	$-$	3	2	2
p_n	1	3	7	17
q_n	0	1	2	5
$p_n^2 + p_n q_n - 15q_n^2$	$-$	-3	3	-1

Concluimos que la unidad fundamental es $\eta = 17 + 5\omega$ y cumple $\mathrm{N}(\eta) = -1$. Sus potencias son

n	1	2	3
η^n	$17 + 5\omega$	$664 + 195\omega$	$25\,913 + 7\,610\omega$

y así vemos que $\epsilon = \eta^3 = 25\,913 + 7\,610\frac{1+\sqrt{61}}{2} = 29\,718 + 3\,805\sqrt{61}$ es la menor potencia de η que está en $\mathcal{O}_2 = \mathbb{Z}\big[\sqrt{61}\,\big]$, por lo que hemos calculado de nuevo la unidad fundamental de este anillo, que ya habíamos calculado para resolver el acertijo del rey Harold. ∎

Ejercicio: Calcular la unidad fundamental del orden $\mathcal{O}_3$ de $\mathbb{Q}(\sqrt{15})$.

10.5 El problema del ganado de Arquímedes

Finalmente estamos en condiciones de resolver el problema del ganado de Arquímedes cuya resolución dejamos inacabada en la introducción. Recordemos que se trata determinar el número de toros y vacas que tiene cada uno de los cuatro rebaños del Sol, y que habíamos llegado a las expresiones

$$\begin{aligned} B &= 46\,200\,808\,287\,018\,u^2 & b &= 32\,116\,937\,723\,640\,u^2\\ N &= 33\,249\,638\,308\,986\,u^2 & n &= 21\,807\,969\,217\,254\,u^2\\ M &= 18\,492\,776\,362\,863\,u^2 & m &= 24\,241\,207\,098\,537\,u^2\\ S &= 32\,793\,026\,546\,940\,u^2 & s &= 15\,669\,127\,269\,180\,u^2 \end{aligned}$$

donde u es un número natural arbitrario, pero falta imponer la condición de que

$$M + S = 51\,285\,802\,909\,803\,u^2$$

sea un número triangular.

Sabemos que un número n es triangular si y sólo si $8n + 1$ es un cuadrado perfecto, luego la condición es que

$$8 \cdot 51\,285\,802\,909\,803\,u^2 + 1 = v^2.$$

Así pues, u tiene que formar parte de una solución de la ecuación de Pell:

$$v^2 - 410\,286\,423\,278\,424u^2 = 1.$$

Se cumple que

$$410\,286\,423\,278\,424 = 2^3 \cdot 3 \cdot 7 \cdot 11 \cdot 29 \cdot 353 \cdot 4\,657^2 = 4\,729\,494 \cdot (2 \cdot 4\,657)^2,$$

por lo que resulta más conveniente buscar una solución de la ecuación

$$v^2 - 4\,729\,494\,w^2 = 1$$

que cumpla además que $2 \cdot 4\,657 = 9\,314 \mid w$ (y entonces $u = w/9\,314$).

En realidad, w es necesariamente par, ya que v tiene que ser impar, luego cumple $v^2 \equiv 1 \pmod 8$, luego $8 \mid 4\,729\,494w^2$, luego $4 \mid 2\,364\,747w^2$, luego $2 \mid w$. Por lo tanto, sólo hace falta exigir que $4\,657 \mid w$.

Para resolver la ecuación de Pell necesitamos

$$\sqrt{4\,729\,494} = [2\,174, \overline{1, 2, 1, 5, 2, 25, 3, 1, 1, 1, 1, 1, 1, 15, 1, 2, 16, 1, 2, 1, 1, 8, 6, 1,}$$

$$\overline{21, 1, 1, 3, 1, 1, 1, 2, 2, 6, 1, 1, 5, 1, 17, 1, 1, 47, 3, 1, 1, 6, 1, 1, 3, 47, 1, 1, 17, 1, 5,}$$

$$\overline{1, 1, 6, 2, 2, 1, 1, 1, 3, 1, 1, 21, 1, 6, 8, 1, 1, 2, 1, 16, 2, 1, 15, 1, 1, 1, 1, 1, 1, 3, 25,}$$

$$\overline{2, 5, 1, 2, 1, 4\,348}]$$

El periodo tiene longitud 92, y es necesario llegar hasta el convergente p_{91}/q_{91} para encontrar la solución fundamental de la ecuación de Pell, que es la asociada a la unidad fundamental:

$$\eta = 109\,931\,986\,732\,829\,734\,979\,866\,232\,821\,433\,543\,901\,088\,049$$

$$+ 50\,549\,485\,234\,315\,033\,074\,477\,819\,735\,540\,408\,986\,340\,\sqrt{4\,729\,494},$$

que cumple, concretamente, $\mathrm{N}(\eta) = 1$, pero no es cierto que $w = 50\,549\ldots$ sea múltiplo de $p = 4\,657$. Necesitamos una potencia $\eta^d = v_d + w_d\sqrt{4\,729\,494}$ que satisfaga $w_d \equiv 0$ (mód p).

Sea $K = \mathbb{Q}\big(\sqrt{\omega}\,\big)$, donde $\omega = \sqrt{4\,729\,494}$. Vamos a ver que basta calcular las potencias de η módulo p, es decir, que podemos trabajar en el anillo k de clases restos módulo p. En primer lugar observamos que

$$\chi_K(p) = \left(\frac{4\,729\,494}{p}\right) = \left(\frac{2\,639}{4\,647}\right) = \left(\frac{2\,647}{2\,639}\right) = \left(\frac{-621}{2\,639}\right) =$$

$$-\left(\frac{2\,639}{621}\right) = -\left(\frac{155}{621}\right) = -\left(\frac{621}{155}\right) = -\left(\frac{1}{155}\right) = -1,$$

luego el teorema 9.16 nos da que el anillo de clases restos módulo p es un cuerpo k con p^2 elementos, que se expresan de forma única como $[a]+[b][\omega]$, para ciertos $0 \le a, b < p$. (Usamos la notación $[\alpha]$ para la clase módulo p para que no se confunda con el conjugado $\bar\alpha$ en K.)

Por lo tanto, si $\eta^d = v_d + w_d\,\omega$, se cumplirá que $w_d \equiv 0$ (mód p) si y sólo si $[\eta]^d = [v_d] + [w_d][\omega] = [v_d]$, es decir, si y sólo si $[\eta]^d$ está en $\mathbb{Z}_p$. (Notemos que aquí usamos la unicidad de la expresión, pues si $p \nmid w_d$, entonces tenemos que $[\eta^d] = [v_d] + [w_d][\omega] \ne [0] + [0][\omega] = [0]$.)

En estos términos, buscamos el menor exponente $d \ge 1$ para el que $[\eta]^d$ está en $\mathbb{Z}_p$.

Observemos que si $l \ge 1$ es cualquier otro entero que cumpla esta misma propiedad, se cumple que $d \mid l$, pues en principio $l = dc + r$, con $0 \le r < d$, pero entonces $[\eta]^l = [\eta^d]^c[\eta]^r$, luego $[\eta]^r = [\eta]^l[\bar\eta^d]^{-c}$ está en $\mathbb{Z}_p$ y, por la minimalidad de d, tiene que ser $r = 0$, luego $d \mid r$.

Esto es útil porque ahora vamos a probar que $[\eta]^{p+1} = [1]$, lo que nos dará que $d \mid p+1$.

En efecto, si $\eta = v_1 + w_1\omega$, tenemos que

$$[\eta]^p = [v_1]^p + [w_1]^p[\omega]^p = [v_1] + [w_1][4\,729\,494]^{(p-1)/2}[\omega] = [v_1] - [w_1][\omega] = [\bar\eta],$$

donde hemos usado que todo elemento de $\mathbb{Z}_p$ cumple $x^p = x$, así como el criterio de Euler, 5.8, junto con que

$$\left(\frac{4\,729\,494}{p}\right) = -1.$$

Por consiguiente, $[\eta]^{p+1} = [\eta][\bar\eta] = [1]$.

Así pues, ahora sabemos que $d \mid p + 1 = 4\,658 = 2 \cdot 17 \cdot 137$, lo cual no deja muchas posibilidades. Más aún, si llamamos $m = 17 \cdot 137 = 2\,329$, hemos probado que $[\eta]^{2m} = [1]$, luego, $[\eta]^m = \pm[1]$ (aquí usamos que k es un cuerpo y que, por lo tanto, el polinomio $x^2 - 1$ sólo tiene las raíces ± 1). Por consiguiente, podemos afirmar que $d \mid m$.

En este punto ya tenemos que calcular potencias de $[\eta]$, para lo cual observamos que

$$[\eta] = [4\,406] + [3\,051][\omega] = -[251] - [1\,606][\omega]$$

así como que $[\omega]^2 = [4\,729\,494] = [2\,639]$.

La tabla siguiente muestra el resultado de los cálculos que prueban que $[\eta]^{17}$ y $[\eta]^{137}$ no están en $\mathbb{Z}_p$:

η	η^2	η^4	η^8	η^{16}	η^{17}
$-251 - 1\,606\omega$	$262 + 551\omega$	$2\,234 + 2\,234\omega$	$1\,560 + 1\,890\omega$	$634 + 1\,038\omega$	$-1\,411 + 1\,933\omega$
η^{32}	η^{64}	η^{128}	η^{136}	η^{137}	
$-1\,750 - 1\,747\omega$	$1\,044 - 141\omega$	$395 - 1\,017\omega$	$2\,603 - 1\,710\omega$	$1\,686 + 2\,334\omega$	

Por lo tanto, concluimos que $d = 2\,329$ y cualquier potencia η^{dr}, para $r = 1, 2, \ldots$ es de la forma $\eta^{dr} = v_{dr} + w_{dr}\omega$ con $9\,314 \mid w_{dr}$. Explícitamente,

$$w_{dr} = \frac{\eta^{dr} - \bar{\eta}^{dr}}{2\sqrt{4\,729\,494}},$$

luego

$$u_r = \frac{\eta^{dr} - \bar{\eta}^{dr}}{2 \cdot 9\,314\sqrt{4\,729\,494}}$$

y

$$u_r^2 = \frac{(\eta^{dr} - \bar{\eta}^{dr})^2}{1\,641\,145\,693\,113\,696} = \frac{\eta^{4\,658r} + \bar{\eta}^{4\,658r} - 2}{1\,641\,145\,693\,113\,696}.$$

Si sustituimos este valor de u^2 en la expresión que teníamos para la solución de problema, resulta

$$B = \tfrac{159}{5\,648}\,\alpha_r, \quad N = \tfrac{801}{39\,536}\,\alpha_r, \quad M = \tfrac{891}{79\,072}\,\alpha_r, \quad S = \tfrac{395}{19\,768}\,\alpha_r,$$
$$b = \tfrac{128\,685}{6\,575\,684}\,\alpha_r, \quad n = \tfrac{2\,446\,623}{184\,119\,152}\,\alpha_r, \quad m = \tfrac{5\,439\,213}{368\,238\,304}\,\alpha_r, \quad s = \tfrac{125\,565}{13\,151\,368}\,\alpha_r,$$

donde $\alpha_r = \eta^{4\,658r} + \bar{\eta}^{4\,658r} - 2$. El número total de toros y vacas es

$$T = \tfrac{25\,194\,541}{184\,119\,152}\,\alpha_r.$$

Ahora bien,

$$0 < \bar{\eta}^{4\,658r} = \frac{1}{\eta^{4\,658r}} < 1,$$

luego

$$-1 < \bar{\eta}^{4\,658r} - 2 < 0$$

luego

$$-1 < \frac{159}{5\,648}(\bar{\eta}^{4\,658r} - 2) < 0,$$

luego

$$B = \left\lceil \frac{159}{5\,648}\eta^{4\,658r} \right\rceil,$$

donde $\lceil\alpha\rceil$ representa el menor entero mayor que α. Lo mismo vale para las demás variables, con lo que una expresión para las soluciones del problema del ganado es:

$$\begin{aligned}
B &= \left\lceil \frac{159}{5\,648}\,\eta^{4\,658r} \right\rceil, \quad & b &= \left\lceil \frac{128\,685}{6\,575\,684}\,\eta^{4\,658r} \right\rceil,\\
N &= \left\lceil \frac{801}{39\,536}\,\eta^{4\,658r} \right\rceil, & n &= \left\lceil \frac{2\,446\,623}{184\,119\,152}\,\eta^{4\,658r} \right\rceil,\\
M &= \left\lceil \frac{891}{79072}\,\eta^{4\,658r} \right\rceil, & m &= \left\lceil \frac{5\,439\,213}{368\,238\,304}\,\eta^{4\,658r} \right\rceil,\\
S &= \left\lceil \frac{395}{19\,768}\,\eta^{4\,658r} \right\rceil, & s &= \left\lceil \frac{125\,565}{13\,151\,368}\,\eta^{4\,658r} \right\rceil,
\end{aligned}$$

para $r = 1, 2, 3, \ldots$, y el número total de animales es

$$T = \left\lceil \frac{25\,194\,541}{184\,119\,152}\,\eta^{4\,658r} \right\rceil$$

Teniendo en cuenta que $\eta + \bar{\eta}$ es un número natural y que $\bar{\eta} = 1/\eta > 0$ es muy pequeño, resulta que η es casi un número natural. Concretamente,

$$\eta \approx 219\,863\,973\,465\,659\,469\,959\,732\,465\,642\,867\,087\,802\,176\,097+$$

$$0.999999999999999999999999999999999999999999954517332502\ldots$$

por lo que

$$\eta \approx 219\,863\,973\,465\,659\,469\,959\,732\,465\,642\,867\,087\,802\,176\,098$$

es una aproximación muy buena. Si calculamos T con tanto con valor exacto como con el aproximado (para $r = 1$) obtenemos que

$$T = 7.760271406486818269530232833213886664232322405923376$$

$$10315061922690321593061406953 1\ldots \times 10^{206\,544}$$

Nota No necesitamos aproximar η, pues *Mathematica* calcula el valor exacto en un tiempo inapreciable con el código obvio:

```
Eta = 109931986732829734979866232821433543901088049 +
50549485234315033074477819735540408986340 Sqrt[4729494];
Round[25194541/184119152 Eta^4658]
```

Las últimas cifras que proporciona para T son

$$\ldots 500\,354\,883\,118\,973\,723\,406\,626\,719\,455\,081\,800.$$

Podemos comprobar que el cálculo es correcto calculando el resto de T módulo 10^k para algún valor razonable de k. Para ello observamos que, si $\eta^{4658} = a + b\omega$, entonces

$$\alpha_1 = \eta^{4658} + \bar{\eta}^{4658} - 2 = 2a - 2,$$

y

$$T = \frac{25\,194\,541}{184\,119\,152}\,\alpha_r = \frac{25\,194\,541}{11\,507\,447}\,\frac{2a-2}{16} = \frac{25\,194\,541}{11\,507\,447}\,\frac{a-1}{8},$$

donde hemos separado el factor 16 del denominador para que éste sea primo con 10 y tenga inverso módulo 10^k. Así, si calculamos $\alpha_1 = u + v\omega$ (mód $8 \cdot 10^k$), tendremos que $a \equiv u$ (mód $8 \cdot 10^k$), luego

$$\frac{a-1}{8} \equiv \frac{u-1}{8} \text{ (mód } 10^k),$$

y

$$T \equiv \frac{25\,194\,541}{11\,507\,447}\,\frac{u-1}{8} \text{ (mód } 10^k).$$

Por ejemplo, si tomamos $k = 6$, podemos calcular los restos módulo $8 \cdot 10^6$ de las potencias de η:

η	$5088049 + 986340\omega$	η^2	$5252801 + 4501320\omega$	η^4	$4691201 + 4394640\omega$
η^8	$5644801 + 7125280\omega$	η^{16}	$4659201 + 7338560\omega$	η^{32}	$3916801 + 4181120\omega$
η^{64}	$4147201 + 5994240\omega$	η^{128}	$268801 + 4244480\omega$	η^{256}	$3955201 + 936960\omega$
η^{512}	$5900801 + 2257920\omega$	η^{1024}	$883201 + 1187840\omega$	η^{2048}	$12801 + 2951680\omega$
η^{4096}	$7731201 + 911360\omega$				

Entonces $\eta^{4658} = \eta^{4096}\eta^{512}\eta^{32}\eta^{16}\eta^2 = 6724801 + 5494280\omega$, luego

$$a \equiv 6\,724\,801 \text{ (mód } 8 \cdot 10^6),$$

luego

$$\frac{a-1}{8} \equiv 840\,600 \text{ (mód } 10^6).$$

Finalmente:

$$25\,194\,541 \equiv 194\,541 \text{ (mód } 10^6), \quad 11\,507\,447^{-1} \equiv 827\,783 \text{ (mód } 10^6),$$

de donde

$$T \equiv 194\,541 \cdot 827\,783 \cdot 840\,600 \equiv 81800 \text{ (mód } 10^6)$$

lo que prueba que las seis últimas cifras de T son 081 800, a lo que hemos llegado con cálculos que no requieren tratar con números mayores que 10^{13}.

Es evidente que Arquímedes no pudo calcular la solución de su problema. Lo que es discutible es si sabía justificar que el problema tiene solución. ∎

10.6 La conjetura de Catalan

En la página 188 vimos que la única solución entera de la ecuación

$$x^p - y^2 = 1,$$

con $p \geq 2$, es $(x, y) = (1, 0)$, mientras que en la página 118 vimos que las únicas soluciones enteras de la ecuación

$$x^2 - y^3 = 1$$

son $(x, y) = (0, -1)$, $(\pm 1, 0)$ y $(\pm 3, 2)$. Ahora vamos a demostrar que éstas son, de hecho, las únicas soluciones enteras de la ecuación

$$x^2 - y^q = 1,$$

con $q \geq 2$. Necesitamos un resultado previo:

Teorema 10.20 *Sean p y q primos no ambos iguales a 2 y sean a, b enteros primos entre sí tales que $a^q - b^q$ es una potencia p-ésima no divisible entre q. Entonces $a - b$ es una potencia p-ésima.*

Demostración: Por hipótesis existe un entero c no divislble entre q de modo que

$$a^q - b^q = (a - b)\,\frac{a^q - b^q}{a - b} = c^p.$$

Vamos a probar que los dos factores son primos entre sí. Para ello probamos que si un primo r cumple $r \mid a - b$, entonces no divide al segundo factor. Notemos que, como $a^q - b^q \equiv a - b \pmod q$ y $q \nmid a^q - b^q$, tiene que ser $q \nmid a - b$, luego $r \neq q$. Si $r \neq 2$ o bien $r = 2$ y $4 \mid a - b$, el teorema 3.6 nos da que

$$v_r\left(\frac{a^q - b^q}{a - b}\right) = v_r(q) = 0.$$

Si $r = 2$ y $a - b \equiv 2 \pmod 4$, entonces, a y b son ambos impares (porque no pueden ser ambos pares, ya que son primos entre sí) y, como $a^q \equiv a \pmod 4$ y $b^q \equiv b \pmod 4$, también $a^q - b^q \equiv 2 \pmod 4$, luego $v_2(a^q - b^q) = 1 = v_2(a - b)$ y llegamos a la misma conclusión.

Si $p \geq 3$, el hecho de que los dos factores sean primos entre sí y el producto sea una potencia p-ésima implica que $a - b$ también es una potencia p-ésima. Podemos suponer, pues, que $p = 2$, con lo que $q \geq 3$.

Ahora sólo podemos asegurar que $a - b = \pm d^2$, para cierto entero d, y tenemos que ver que el signo negativo es imposible. En efecto, en tal caso

$$\frac{a^q - b^q}{a - b} = -\frac{c^2}{d^2} < 0,$$

pero eso no puede ser, pues entonces $a^q - b^q$ y $a - b$ tendrían signos opuestos, lo cual es imposible, porque la función $x \mapsto x^q$ es creciente. ∎

Teorema 10.21 *Las únicas soluciones enteras de la ecuación* $x^2 - y^q = 1$ *con* $q \geq 2$ *son* $(x, y) = (0, -1)$, $(\pm 1, 0)$ *y* $(\pm 3, 2)$.

DEMOSTRACIÓN: Es obvio que si $q = 2$ las únicas soluciones son $(\pm 1, 0)$, pues 0 y 1 son los únicos cuadrados consecutivos. Sea q el menor exponente para el que la ecuación tiene una solución distinta de las indicadas. Entonces q tiene que ser primo, pues si $q = q_1 q_2$, donde $1 < q_1, q_2 < q$, tendríamos que (x, y^{q_2}) sería una solución de la ecuación con exponente q_1, luego $y^{q_2} = -1, 0, 2$, luego $y = -1, 0$ y entonces (x, y) es necesariamente una de las soluciones indicadas.

Como el caso $q = 3$ está resuelto en la página 118 y acabamos de considerar el caso $q = 2$, podemos suponer que $q \geq 5$, y basta probar que las únicas soluciones de la ecuación son $(\pm 1, 0)$ y $(0, -1)$. Ésta puede escribirse en la forma

$$(x + 1)(x - 1) = y^q.$$

Un divisor primo común de $x + 1$ y $x - 1$ divide a su diferencia, luego tiene que ser 2. Distinguimos dos casos:

Si x es par, podemos suponer que $x \geq 2$ (pues $x = 0$ lleva a $y = -1$) y entonces $x + 1$ y $x - 1$ son impares, luego primos entre sí. La ecuación implica entonces que $x + 1$ y $x - 1$ son potencias q-ésimas que difieren en dos unidades, y esto no puede ser, pues

$$(u + 1)^q \geq 1 + qu \geq 6.$$

Por lo tanto, podemos suponer que $x \geq 3$ es impar, y a su vez que $y \geq 2$ es par. Entonces $x + 1$ y $x - 1$ son pares, pero sólo uno de ellos será múltiplo de 4, luego $(x + 1, x - 1) = 2$. Más precisamente, si $x \equiv r \pmod 4$, donde $r = \pm 1$, tenemos que $x + r = 2u$, $x - r = 2^i v$, donde u y v son impares y $(u, v) = 1$, y la ecuación nos da que

$$2^{i+1} uv = y^q,$$

de donde se sigue u y v son potencias q-ésimas. Más precisamente,

$$y = 2ab, \qquad x + r = 2a^q, \quad x - r = 2^{q-1} b^q,$$

donde $(a, b) = 1$ y a es impar. Como $q \geq 5$ y $x \geq 2$, tenemos que

$$\frac{a^q}{b^q} = 2^{q-2} \frac{x + r}{x - r} \geq 8 \frac{x - 1}{x + 1} \geq 2.$$

En particular $a > b$. Un simple cálculo muestra que

$$a^{2q} - (2rb)^q = \left(\frac{x + r}{2}\right)^2 - 2r(x - r) = \left(\frac{x - 3r}{2}\right)^2.$$

El teorema anterior nos da que si $q \nmid (x - 3r)/2$, entonces $a^2 - 2rb$ es un cuadrado, pero esto no es posible, pues

$$(a-1)^2 < a^2 - 2(a-1) \leq a^2 - 2b \leq a^2 - 2rb \leq a^2 + 2b < a^2 + 2a + 1 = (a+1)^2.$$

Concluimos que $q \mid (x-3r)/2$ y en particular $q \nmid x$, pues en caso contrario $q \mid 3r = \pm 3$, pero $q \geq 5$. Por otra parte, como

$$x^2 = y^q - (-1)^q,$$

el teorema anterior nos da que $y + 1 = y - (-1) = k^2$, luego k es impar e $y = k^2 - 1 \geq 2$ no es un cuadrado. Por el teorema 10.18, la solución fundamental de la ecuación de Pell

$$u^2 - yv^2 = 1$$

es $(k, 1)$. Como otra solución es $(x, y^{(q-1)/2})$, tiene que existir un m tal que

$$x + y^{(q-1)/2}\sqrt{y} = (k + \sqrt{y})^m,$$

luego

$$x \equiv k^m + mk^{m-1}\sqrt{y} \text{ (mód } y),$$

luego $y \mid k^m - x$, $y \mid mk^{m-1}$. Como y es par y k impar, resulta que m es par. Por otra parte,

$$x + y^{(q-1)/2}\sqrt{y} \equiv (\sqrt{y})^m \equiv y^{m/2} \text{ (mód } k),$$

luego $k \mid x - y^{m/2}$, $k \mid y^{(q-1)/2}$.

Si es $k \geq 2$, podemos tomar un divisor primo $r \mid k$, y a su vez $r \mid y$, luego $r \mid k^2 - y = 1$, y tenemos una contradicción. Por lo tanto tiene que ser $k = 1$, luego $y = 0$, contradicción. ∎

Recordemos (véase la introducción) que la conjetura de Catalan afirma que la ecuación

$$x^p - y^q = 1$$

no tiene soluciones enteras no triviales (es decir, con $x, y \geq 1$, $p, q \geq 2$) salvo $3^2 - 2^3 = 1$. Con lo que hemos probado es inmediato el teorema siguiente:

Teorema 10.22 *La conjetura de Catalan es equivalente a que si p y q son primos impares, entonces la ecuación*

$$x^p - y^q = 1$$

no tiene soluciones enteras con $x, y \geq 1$.

Demostración: Ciertamente, el enunciado es un caso particular de la conjetura de Catalan. Supongamos la afirmación del enunciado y consideremos números $m, n \geq 2$ y $x, y \geq 1$ de modo que $x^m - y^n = 1$.

Si m es par, entonces $(x^{m/2})^2 - y^n = 1$, luego el teorema anterior nos da que $(x^{m/2}, y) = (3, 2)$, luego $m = 2$ y $(x, y) = (3, 2)$. Por lo tanto, podemos suponer que m es impar. Similarmente, si n es par $x^m - (y^{n/2})^2 = 1$ contradice el ejemplo de la página 188, luego n tiene que ser impar. Ahora tomamos divisores primos $p \mid m$ y $q \mid n$, de modo que $(x^{m/p})^p - (y^{n/q})^q = 1$ contradice la hipótesis. ∎

Sucede que ésta es la formulación más adecuada de la conjetura de Catalan para proceder a su demostración. El nivel de la prueba excede con mucho las posibilidades de este libro, pero observemos que la hemos demostrado para $p = 2$ y para $q = 2$. Por otro lado, el teorema de Gersónides 3.4 es otro caso particular, esta vez para $(x, y) = (2, 3)$, $(3, 2)$ y exponentes arbitrarios. Ahora podemos generalizarlo sustancialmente. El teorema siguiente generaliza el caso $(2, 3)$:

Teorema 10.23 *Si $x \geq 2$ es potencia de 2, la ecuación $x^m - y^n = 1$ no tiene soluciones enteras con $x, y \geq 1$, $m, n \geq 2$.*

DEMOSTRACIÓN: Como en la prueba precedente, si n es par resulta que $x^m - (y^{n/2})^2 = 1$ contradice el ejemplo de la página 188, luego n tiene que ser impar y podemos tomar un primo (impar) $q \mid n$ tal que la ecuación $x^m - y^q = 1$ tiene solución. Entonces

$$x^p = 1 + y^q = (1 + y)(1 - y + y^2 - \cdots + y^{q-1}).$$

Como x es potencia de 2, el factor $1 + y$ también lo es, luego y es impar, luego en el factor derecho hay q sumandos impares, luego es impar, y tenemos una contradicción. ∎

El teorema siguiente generaliza el caso $(3, 2)$:

Teorema 10.24 *Si $x \equiv 3, 5, 7 \pmod 8$, la ecuación $x^m - y^n = 1$ no tiene soluciones enteras con $x, y \geq 1$, $m, n \geq 2$ excepto $3^2 - 2^3 = 1$.*

DEMOSTRACIÓN: Si m es par, entonces $(x^{m/2})^2 - y^n = 1$, luego 10.21 prueba que $(x^{m/2}, y) = (3, 2)$, luego $m = 2$, luego $(x, y) = (3, 2)$. Por lo tanto, podemos suponer que m es impar. Similarmente, si n es par tenemos que $x^m - (y^{m/2})^2 = 1$ contradice el ejemplo de la página 188. Ahora tomamos primos impares $p \mid m$, $q \mid n$ de modo que $(x^{m/p})^p - (y^{n/q})^q = 1$. Más aún, si x y k son impares, es fácil ver que $x^k \equiv x \pmod 8$, luego la ecuación

$$x^p - y^q = 1$$

también tiene una solución con $x \equiv 3, 5, 7 \pmod 8$. Como x es impar, vemos que y es par y, como $q \geq 3$, resulta que $x^p - 1 = y^q \equiv 0 \pmod 8$. Sin embargo, esto es imposible, pues

$$x^p - 1 = x(x^{p-1} - 1) + x - 1 = x(x^{(p-1)/2} + 1)(x^{(p-1)/2} - 1) + x - 1$$

y $x^{(p-1)/2} \pm 1$ son dos números pares consecutivos, luego uno de ellos es múltiplo de 4 y el producto es múltiplo de 8. Por lo tanto, $x^p - 1 \equiv x - 1 \not\equiv 0 \pmod 8$. ∎

El teorema siguiente generaliza al teorema de Gersónides completo:

Teorema 10.25 *Si y es potencia de primo, la ecuación $x^p - y^q = 1$ no tiene soluciones enteras $x, y \geq 1$, $p, q \geq 2$ excepto $3^2 - 2^3 = 1$.*

DEMOSTRACIÓN: Como en los casos precedentes, podemos suponer que p y q son primos impares (ahora usamos que una potencia de una potencia de primo es potencia de primo). La ecuación equivale a

$$y^q = (x-1)\frac{x^p-1}{x-1}$$

Hemos probado que la ecuación no tiene solución cuando $x = 2, 3$, luego el primer factor es $x - 1 > 1$. Por otro lado,

$$\frac{x^p-1}{x-1} = x^{p-1} + x^{p-2} + \cdots + x + 1 \equiv (p-1)x + 1 \text{ (mód 2)},$$

lo que prueba que el segundo factor es impar. Más aún, la primera igualdad prueba que es mayor que $x + 1$. En particular, si $y = r^k$, con r primo, resulta que r es impar. Pongamos que

$$x - 1 = r^a, \qquad \frac{x^p-1}{x-1} = r^b,$$

donde $1 \le a < b$ y $a + b = kq$. El teorema 3.6 implica que

$$b = v_r\left(\frac{x^p-1}{x-1}\right) = v_r(p).$$

por lo tanto, $p = r$ y $b = 1$, con lo que tenemos una contradicción. ∎

Capítulo XI

Formas cuadráticas

En este capítulo vamos a iniciar el estudio general de las ecuaciones diofánticas de la forma

$$ax^2 + bxy + cy^2 = m,$$

donde a, b, c son números enteros. Ya hemos estudiado algunos casos particulares de estas ecuaciones. Por ejemplo, sabemos que la ecuación de Pell $x^2 - Dy^2 = 1$ tiene infinitas soluciones cuando $D > 0$ no es un cuadrado perfecto, mientras que con $m = -1$ puede no tener solución. También sabemos para qué valores de m la ecuación $x^2 + y^2 = m$ tiene solución, etc. Ahora vamos a exponer una teoría general para estas ecuaciones, si bien un algoritmo de resolución tendrá que esperar hasta el capítulo siguiente.

11.1 Conceptos básicos sobre formas cuadráticas

Los polinomios de la forma $ax^2+bxy+cy^2$ se llaman *formas cuadráticas* (de dos variables). En principio, los coeficientes pueden estar en cualquier anillo, pero aquí vamos a considerar únicamente el caso de que sean números racionales y, especialmente, enteros.

Diremos que un entero r *está representado* por una forma cuadrática $f(x, y)$ si existen enteros x, y tales que $f(x, y) = r$.

Formas primitivas Una forma cuadrática $f(x, y) = ax^2 + bxy + cy^2$ con coeficientes enteros no todos nulos se dice *primitiva* si sus coeficientes son primos entre sí (es decir, que ningún primo divide a los tres, pero sí que puede dividir a dos de ellos).

En general, si llamamos d al máximo común divisor de a, b y c, podemos descomponer

$$ax^2 + bxy + cy^2 = d(a'x^2 + b'xy + c'y^2),$$

donde $a' = a/d$, $b' = b/d$ y $c' = c/d$, y la forma $f'(x, y) = a'x^2 + b'xy + c'y^2$ es primitiva. Es claro entonces que para que una ecuación diofántica $f(x, y) = m$

pueda tener solución es necesario que $d \mid m$, y en tal caso las soluciones serán las mismas que las de la ecuación $f'(x,y) = m/d$.

Por lo tanto, a la hora de resolver ecuaciones diofánticas definidas por formas cuadráticas, no perdemos generalidad si consideramos únicamente formas primitivas.

Formas irreducibles Una forma cuadrática con coeficientes enteros (resp. racionales) se dice *reducible* si puede descomponerse en factores lineales:

$$ax^2 + bxy + cy^2 = (rx + sy)(tx + uy),$$

donde r, s, t, u son números enteros (resp. racionales).

Si una forma cuadrática con coeficientes enteros es reducible, a la hora de resolver una ecuación diofántica

$$(rx + sy)(tx + uy) = m,$$

basta considerar todas las factorizaciones posibles $m = ab$ (que son un número finito) y resolver los sistemas de ecuaciones

$$\begin{aligned} rx + sy &= a \\ tx + uy &= b \end{aligned}$$

(más adelante veremos cómo hacerlo). En cambio, esto no es posible cuando la forma cuadrática es irreducible.

En principio, dada una forma cuadrática con coeficientes enteros, podría ocurrir que fuera irreducible como tal, es decir, como forma con coeficientes enteros, pero que fuera reducible si la consideramos con coeficientes racionales, es decir, podría ocurrir que admitiera una descomposición en producto de dos factores lineales con coeficientes racionales, pero no enteros. Sin embargo, lo cierto es que no puede darse el caso:

> *Si una forma cuadrática con coeficientes enteros factoriza como*
>
> $$ax^2 + bxy + cy^2 = (rx + sy)(tx + uy),$$
>
> *donde los factores tienen coeficientes racionales, entonces éstos pueden tomarse con coeficientes enteros.*

En efecto, si $c = 0$ es obvio que $ax^2 + bxy = x(ax + by)$ es una factorización con coeficientes enteros. Supongamos, pues que $c = su \neq 0$. Entonces

$$ax^2 + bxy + cy^2 = c\left(\frac{r}{s}x + y\right)\left(\frac{t}{u}x + y\right),$$

donde en principio r/s y t/u son cocientes de números racionales, pero operándolos podemos llegar a una expresión idéntica en la que r, s, t, u sean números enteros con $(r, s) = (t, u) = 1$. A su vez, de aquí llegamos a una factorización

$$ax^2 + bxy + cy^2 = \frac{c}{su}(rx + sy)(tx + uy),$$

donde ahora cada letra representa un número entero.

Llamemos d al máximo común divisor de a, b, c, de modo que

$$sud(a'x^2 + b'xy + c'y^2) = c(rx + sy)(tx + uy),$$

donde a', b', c' son primos entre sí.

Basta probar que los coeficientes de

$$(rx + sy)(tx + uy) = rtx^2 + (ru + st)xy + suy^2$$

son primos entre sí, pues entonces, tanto sud como c serán el máximo común divisor de los coeficientes de una misma forma cuadrática, luego será $sud = \pm c$, luego podremos concluir que

$$ax^2 + bxy + cy^2 = \pm d(rx + sy)(tx + uy).$$

En efecto, supongamos que un primo p divide a los tres coeficientes del producto. En particular $p \mid rt$ y, por simetría, podemos suponer que $p \mid r$. Entonces, $p \nmid s$, ya que $(r, s) = 1$. Como $p \mid (ru + st)$, también $p \mid st$, luego $p \mid t$ y como $p \mid su$, también $p \mid u$, luego $p \mid (t, u) = 1$, contradicción. ∎

Esto nos da un criterio muy simple de irreducibilidad:

Definición 11.1 Llamaremos *discriminante* de una forma cuadrática

$$ax^2 + bxy + cy^2$$

al número $D = b^2 - 4ac$.

Teorema 11.2 *Una forma cuadrática $ax^2 + bxy + cy^2$ con coeficientes enteros es irreducible si y sólo si su discriminante no es un cuadrado perfecto.*

Demostración: Acabamos de ver que para que una forma cuadrática sea irreducible es necesario y suficiente que pueda descomponerse como

$$ax^2 + bxy + cy^2 = (rx + sy)(tx + uy).$$

aunque los coeficientes r, s, t, u sean racionales, lo cual a su vez es equivalente a que

$$ax^2 + bx + c = (rx + s)(tx + u).$$

En efecto, de la primera igualdad se pasa a la segunda haciendo $y = 1$, mientras que de la segunda se pasa a la primera sustituyendo x por x/y y multiplicando ambos miembros por y^2.

Así pues, la forma cuadrática es reducible si y sólo si el polinomio ax^2+bx+c es reducible en $\mathbb{Q}[x]$, lo cual es equivalente a que tenga raíces en $\mathbb{Q}$, y a su vez a que su discriminante D sea un cuadrado perfecto, para que sus raíces sean racionales. ∎

Toda forma cuadrática irreducible cumple $a \neq 0$ y $c \neq 0$. Entonces

$$\begin{aligned} ax^2 + bxy + cy^2 &= \frac{1}{4a}((2ax)^2 + 4abxy + 4acy^2) \\ &= \frac{1}{4a}((2ax + by)^2 - (b^2y^2 + 4acy^2)) \\ &= \frac{1}{4a}((2ax + by)^2 - Dy^2). \end{aligned}$$

A partir de esta expresión es inmediato el teorema siguiente:

Teorema 11.3 *Sea $f(x, y)$ una forma cuadrática con coeficientes enteros irreducible de discriminante D. Entonces:*

1. *Si $D < 0$ y $a > 0$ (o, equivalentemente, $c > 0$), entonces $f(x, y) > 0$ siempre que $(x, y) \neq (0, 0)$.*
2. *Si $D < 0$ y $a < 0$ (o, equivalenemente, $c < 0$), entonces $f(x, y) < 0$ siempre que $(x, y) \neq (0, 0)$.*
3. *Si $D > 0$ entonces $f(x, y) > 0$ para algunos pares (x, y) y $f(x, y) < 0$ para otros pares.*

Definición 11.4 En el primer caso del teorema anterior se dice que la forma cuadrática $f(x, y)$ es *definida positiva,* en el segundo que es *definida negativa* y en el tercero que es *indefinida.*

A la hora de resolver una ecuación diofántica $f(x, y) = d$ donde f tiene discriminante $D < 0$, vemos que para que tenga solución es necesario que d tenga el mismo signo que a (o que c) y, en tal caso, pasando si es preciso a la ecuación $-f(x, y) = -d$, que tiene las mismas soluciones, no perdemos generalidad si suponemos que f es definida positiva.

En resumen:

> *Al efecto de estudiar las ecuaciones diofánticas definidas por formas cuadráticas irreducibles, no perdemos generalidad si las suponemos primitivas y, en caso de que tengan discriminante negativo, podemos suponer también que son definidas positivas.*

Cambios de variables Introducimos ahora una idea fundamental en el estudio de las ecuaciones diofánticas definidas por formas cuadráticas. Supongamos que nos interesan las soluciones enteras de una ecuación como

$$13x^2 + 10xy + 2y^2 = 10. \tag{11.1}$$

En principio no sabemos cómo encontrarlas, pero observemos lo que sucede si hacemos el cambio de variables

$$x = -x' + y', \qquad y = 3x' - 2y'.$$

Un cálculo rutinario muestra que

$$13(-x'+y')^2+10(-x'+y')(3x'-2y')+2(3x'-2y')^2=x'^2+y'^2.$$

Por lo tanto, para cada solución entera (x',y') de la ecuación

$$x'^2+y'^2=10, \tag{11.2}$$

los valores (x,y) que resultan de las ecuaciones de cambio de variable nos dan una solución de la ecuación original. Concretamente, es claro que las únicas soluciones enteras de esta ecuación son $(x',y')=(\pm3,\pm1),(\pm1,\pm3)$. Por lo tanto,

$$(x,y)=(-2,7),(2,-7),(4,-11),(-4,11),(2,-3),(-2,3),(4,-9),(-4,9)$$

son ocho soluciones enteras de la ecuación original. ¿Són las únicas? Sí. La razón es que en las ecuaciones del cambio de variables podemos despejar las variables x' e y': en la primera despejamos $x'=-x+y'$ y, al sustituir en la segunda, queda

$$y=3(-x+y')-2y'=-3x+3y'-2y'=-3x+y',$$

luego $y'=3x+y$, y a su vez $x'=-x+3x+y=2x+y$. Esto se traduce en que los cambios de variable

$$\left.\begin{array}{l} x=-x'+\ y' \\ y=\ 3x'-2y' \end{array}\right\} \qquad \left.\begin{array}{l} x'=2x+y \\ y'=3x+y \end{array}\right\} \tag{11.3}$$

son mutuamente inversos. Si sustituimos el segundo cambio en (11.2) obtenemos la ecuación (11.1). Por lo tanto, si (x,y) es una solución entera de (11.1) entonces el valor (x',y') dado por el segundo cambio de variables será una de las 8 soluciones de la ecuación (11.2), de donde se sigue a su vez que (x,y) será una de las 8 soluciones que hemos encontrado.

Ejercicio: Demostrar que la ecuación $13x^2+10xy+2y^2=7$ no tiene soluciones enteras.

Con esto hemos "resuelto" la ecuación diofántica (11.1), en el sentido de que hemos encontrado todas sus soluciones enteras, pero no podemos decir que dispongamos de un método de resolución, pues no sabemos por qué tendríamos que haber considerado precisamente el cambio de variables que nos ha permitido encontrarlas. Además, el cambio de variables que hemos encontrado "no es cualquiera ", como ilustra el ejemplo siguiente:

Consideremos ahora la ecuación

$$19x^2+15xy+3y^2=61.$$

Si se nos ocurre aplicarle el cambio de variables

$$x=x'+3y', \qquad y=-3x'-7y' \tag{11.4}$$

se transforma en otra más sencilla que ya hemos estudiado:

$$x'^2 + 3y'^2 = 61.$$

Es fácil concluir que tiene exactamente cuatro soluciones: $(x', y') = (\pm 7, \pm 2)$, que, a través del cambio de variables, se corresponden con las soluciones

$$(x, y) = (-1, 7), \quad (1, -7), \quad (13, -35), \quad (-13, 35).$$

¿Son las únicas? En este caso no. Las soluciones enteras de la ecuación original son estas cuatro más otras ocho:

$$(x, y) = (1, 2), \quad (-1, -2), \quad (13, -30), \quad (-13, 30),$$
$$(14, -33), \quad (14, -37), \quad (-14, -33), \quad (-14, 37).$$

¿Y por qué hemos perdido soluciones en este caso? La explicación es que si calculamos el cambio de variable inverso, éste resulta ser:

$$x' = \frac{1}{2}(-7x - 3y), \qquad y' = \frac{1}{2}(3x + y).$$

Las soluciones que hemos encontrado son las soluciones (x, y) que tienen ambas coordenadas impares, lo cual se traduce en que (x', y') son enteros y, por lo tanto, son soluciones enteras de la ecuación $x'^2 + 3y'^2 = 61$. En cambio, las soluciones (x, y) con una coordenada par y otra impar se corresponden con soluciones fraccionarias, no enteras, de la ecuación $x'^2 + 3y'^2 = 61$.

Vemos así que, para que un cambio de variables nos permita reducir el cálculo de las soluciones enteras de una ecuación al cálculo de las de otra, no sólo es necesario que tenga coeficientes enteros, sino que al cambio inverso le ocurra lo mismo, cosa que no se cumplía en este último ejemplo. De hecho, antes que esto es necesario que exista el cambio inverso, pues podría no existir. Por ejemplo, si intentamos hacer el cambio de variables

$$x = 2x' + 2y', \qquad y = 3x' + 3y',$$

observamos que todos los pares (x, y) que proceden de pares (x', y') cumplen $3x = 2y$, luego un par como $(x, y) = (1, 1)$ no procede de ningún par (x', y') por este cambio de variables, luego no hay cambio de variables inverso. Si $(x, y) = (1, 1)$ fuera una solución entera de una ecuación determinada por una forma cuadrática, no podríamos obtenerla a partir de una solución (x', y') de la forma resultante de aplicar este cambio de variables. En la sección siguiente estudiaremos estas cuestiones con detalle.

11.2 Matrices

Aunque nos van a interesar principalmente los casos $A = \mathbb{Z}$ y $A = \mathbb{Q}$, de momento podemos considerar un dominio arbitrario A. Podemos definir un *cambio de variables lineal* (de dos variables) en A como una aplicación

$$f : A^2 \longrightarrow A^2$$

de la forma $f(x', y') = (ax' + cy', bx' + dy')$, para ciertos a, b, c, d en A.

Por ejemplo, las ecuaciones (11.3) definen dos cambios de variables f y g en $\mathbb{Z}^2$. El primero hace corresponder la solución $(x', y') = (3, 1)$ de la ecuación (11.2) con la solución $(x, y) = f(3, 1) = (-2, 7)$ de la ecuación (11.1), mientras que el segundo hace corresponder $(x, y) = (-2, 7)$ con $g(-2, 7) = (3, 1)$.

Dados dos cambios de variables

$$A^2 \xrightarrow{f} A^2 \xrightarrow{g} A^2,$$

podemos considerar su *composición* $f \circ g$, es decir, el cambio de variables dado por

$$(f \circ g)(x', y') = g(f(x', y')).$$

Por ejemplo, la composición de los cambios de variables

$$f \equiv \begin{cases} x = 3x' + y' \\ y = 2x' + 3y' \end{cases} \qquad g \equiv \begin{cases} u = x - y \\ v = -x + 2y \end{cases} \tag{11.5}$$

es el cambio dado por

$$\begin{aligned} (f \circ g)(x', y') &= g(f(x', y')) = g(3x' + y', 2x' + 3y') \\ &= (3x' + y' - (2x' + 3y'), -(3x' + y') + 2(2x' + 3y')) \\ &= (x' - 2y', x' + 5y') \end{aligned}$$

o, equivalentemente,

$$f \circ g \equiv \begin{cases} u = x' - 2y' \\ v = x' + 5y' \end{cases} \tag{11.6}$$

Otro ejemplo: en la sección anterior hemos visto que la composición de los cambios de variables (11.3) es el cambio de variables trivial $u = x'$, $v = y'$, de modo que un cambio "deshace" al otro.

En general, la composición de dos cambios de variables

$$f \equiv \begin{cases} x = ax' + cy' \\ y = bx' + dy' \end{cases} \qquad g \equiv \begin{cases} u = a'x + c'y \\ v = b'x + d'y \end{cases}$$

viene dada por

$$f \circ g \equiv \begin{cases} x = (aa' + bc')x' + (ca' + dc')y' \\ y = (ab' + bd')x' + (cb' + dd')y' \end{cases} \tag{11.7}$$

La composición de cambios de variable es asociativa, es decir, que si tenemos tres cambios de variable

$$A^2 \xrightarrow{f} A^2 \xrightarrow{g} A^2 \xrightarrow{h} A^2,$$

entonces $(f \circ g) \circ h = f \circ (g \circ h)$.

En efecto,

$$((f \circ g) \circ h)(x', y') = h((f \circ g)(x', y')) = h(g(f(x', y'))),$$

$$(f \circ (g \circ h))(x', y') = (g \circ h)(f(x', y')) = h(g(f(x', y'))),$$

luego $((f \circ g) \circ h)(x', y') = (f \circ (g \circ h))(x', y')$.

Todos los ejemplos que hemos considerado hasta ahora son conceptualmente muy simples, pero los cálculos resultan farragosos. Vamos a ver que esto cambia completamente si introducimos una notación matricial.

Un cambio de variables

$$f \equiv \begin{cases} x = ax' + cy' \\ y = bx' + dy' \end{cases}$$

puede "resumirse" en la *matriz*

$$M_f = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Notemos que la primera fila de M_f es $f(1,0)$ y la segunda es $f(0,1)$. A partir de f podemos calcular así M_f y a partir de M_f podemos reconstruir f. Esto significa que podemos pensar en los cambios de variables y en las matrices como dos formas equivalentes de representar un mismo concepto.

En términos de matrices, las fórmulas (11.7) para la composición de dos cambios de variables arbitrarios se expresan en términos del producto de matrices definido como sigue:

$$\begin{pmatrix} a & b \\ \mathbf{c} & \mathbf{d} \end{pmatrix} \begin{pmatrix} \mathbf{a'} & b' \\ \mathbf{c'} & d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ \mathbf{ca' + dc'} & cb' + dd' \end{pmatrix}. \tag{11.8}$$

Esto significa que si dos cambios f y g se corresponden con las matrices M_f y M_g, respectivamente, entonces la composición $f \circ g$ se corresponde con el producto $M_f M_g$ definido por la igualdad anterior.

La ventaja es que, con esta representación, el resultado de la composición puede calcularse sin necesidad de recordar ninguna fórmula. Por ejemplo, para calcular el valor $ca' + dc'$ que hemos destacado en negrita, sólo tenemos que tener en cuenta que es el situado en la segunda fila y en la primera columna de la matriz producto, y vemos que se obtiene a partir de la segunda fila del primer factor y la primera columna del segundo. Sólo hay que recorrer la fila de izquierda a derecha y la columna de arriba a abajo multiplicando los números que encontramos y sumando al pasar a los siguientes. Y lo mismo vale para las demás entradas de la matriz producto. Cada una de ellas se obtiene recorriendo la fila correspondiente del primer factor y la columna correspondiente del segundo.[1]

Por ejemplo, antes hemos calculado la composición de los cambios (11.5), y ahora podemos calcularla mucho más cómodamente multiplicando las matrices correspondientes:

$$\begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -2 & 5 \end{pmatrix} \tag{11.9}$$

vemos que, en efecto, la matriz resultante es la correspondiente al cambio (11.6).

[1] Si el lector no está familiarizado con el producto de matrices en este punto debería ponerse a sí mismo productos arbitrarios hasta que vea que es capaz de calcularlos sin necesidad de recordar ninguna fórmula.

Similarmente, la comprobación de que los cambios (11.3) son mutuamente inversos se reduce ahora a calcular el producto

$$\begin{pmatrix} -1 & 3 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

pues la matriz *identidad*

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{11.10}$$

se corresponde ciertamente con el cambio de variables trivial $f(x, y) = (x, y)$.

Vamos a recapitular los conceptos que acabamos de introducir:

Definición 11.5 Si A es un dominio, llamaremos $\mathrm{Mat}_2(A)$ al conjunto de todas las *matrices* (con dos filas y dos columnas) con coeficientes en A, es decir, todos los conjuntos de cuatro elementos de A dispuestos en la forma

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

En $\mathrm{Mat}_2(A)$ tenemos definido el producto dado por (11.8), que tiene la propiedad asociativa, pues hemos visto que, a través de la identificación entre matrices y cambios de variables se corresponde con la composición, y ésta es asociativa. Además es fácil ver que tiene como elemento neutro a la matriz identidad (11.10).

Es muy importante tener presente que el producto de matrices no es conmutativo. Por ejemplo, basta comparar (11.9) con

$$\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix}.$$

Más delicada es la cuestión de cuándo una matriz tiene una matriz inversa.

Diremos que una matriz M de $\mathrm{Mat}_2(A)$ es *regular* si existe otra matriz M^{-1} en $\mathrm{Mat}_2(A)$ tal que $MM^{-1} = M^{-1}M = I$. En caso contrario se dice que es *singular*.

Enseguida veremos (véase la observación tras el teorema 11.7) que, aunque el producto de matrices no sea conmutativo, si dos matrices cumplen $AB = I$, necesariamente $BA = I$, luego en la definición de matriz regular podemos omitir una de las igualdades.

De momento observemos que la matriz M^{-1}, si existe, es única, y se llama *matriz inversa* de M. En efecto, si M^* fuera otra matriz inversa, tendríamos que

$$M^* = M^*I = M^*MM^{-1} = IM^{-1} = M^{-1}.$$

La discusión de la sección precedente muestra que, si queremos reducir el estudio de una ecuación diofántica determinada por una forma cuadrática al de otra (pretendidamente más simple) obtenida mediante un cambio de variables,

para que el cambio haga corresponder las soluciones de ambas ecuaciones necesitamos que sea inversible, es decir, que se corresponda con una matriz regular. Más precisamente, que sea regular como matriz de $\mathrm{Mat}_2(\mathbb{Z})$, porque hemos visto que una misma matriz puede ser singular en $\mathrm{Mat}_2(\mathbb{Z})$ y regular en $\mathrm{Mat}_2(\mathbb{Q})$. Es el caso de

$$M = \begin{pmatrix} 1 & -3 \\ 3 & -7 \end{pmatrix} \tag{11.11}$$

que se corresponde con el cambio de variables (11.4). Los cálculos que hemos hecho sobre este cambio de variables equivalen a que

$$\begin{pmatrix} 1 & -3 \\ 3 & -7 \end{pmatrix} \begin{pmatrix} -7/2 & 3/2 \\ -3/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

luego M es regular en $\mathrm{Mat}_2(\mathbb{Q})$, pero no puede serlo en $\mathrm{Mat}_2(\mathbb{Z})$, pues una inversa de M en $\mathrm{Mat}_2(\mathbb{Z})$ lo sería también en $\mathrm{Mat}_2(\mathbb{Q})$, luego tendría que ser la que aparece en la igualdad anterior, que no está en $\mathrm{Mat}_2(\mathbb{Z})$.

Vamos a ver que existe un criterio muy simple para saber si una matriz es regular y, en su caso, para calcular su inversa. Vamos a considerar el problema en términos de cambios de variables. Tenemos un cambio

$$\left.\begin{aligned} x &= ax' + cy' \\ y &= bx' + dy' \end{aligned}\right\}$$

y nos planteamos en qué condiciones existirá un cambio inverso que "deshaga" lo que éste "hace". Para ello aplicamos técnicas usuales en la resolución de ecuaciones. Multiplicamos la primera ecuación por d y la segunda por c. El resultado es

$$\left.\begin{aligned} dx &= adx' + cdy' \\ cy &= bcx' + cdy' \end{aligned}\right\}$$

Restando ambas ecuaciones resulta: $(ad - bc)x' = dx - cy$. Similarmente, si multiplicamos la primera por b, la segunda por a y restamos, llegamos en total a

$$\left.\begin{aligned} (ad - bc)x' &= dx - cy \\ (ad - bc)y' &= -bx + ay \end{aligned}\right\}$$

Por lo tanto, si $ad - bc$ es una unidad del anillo A en el que estamos trabajando, podemos acabar de despejar

$$\left.\begin{aligned} x' &= \tfrac{1}{ad-bc}(dx - cy) \\ y' &= \tfrac{1}{ad-bc}(-bx + ay) \end{aligned}\right\}$$

Y concluimos que existe un cambio de variables inverso. Vamos a expresar esto en términos de matrices:

Definición 11.6 El *determinante* de una matriz

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

se define como $|M| = \det M = ad - bc$.

Es claro que $|I| = 1$, y una simple comprobación muestra que se cumple la relación

$$|MN| = |M||N|$$

(basta calcular los dos miembros para dos matrices arbitrarias y ver que los resultados coinciden).

Teorema 11.7 *Una matriz*

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

con coeficientes en un dominio A es regular en $\mathrm{Mat}_2(A)$ *si y sólo si $|M|$ es una unidad de A. En tal caso la matriz inversa es*

$$M^{-1} = \frac{1}{|M|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

(donde hay que entender que el producto de un elemento d de A por una matriz se calcula multiplicando por d todas las entradas de la matriz).

Demostración: Se comprueba inmediatamente que

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} |M| & 0 \\ 0 & |M| \end{pmatrix},$$

lo que a su vez implica que, si $|M|$ es una unidad, se cumplen las relaciones $MM^{-1} = M^{-1}M = I$, luego M es regular.

Recíprocamente, si M es regular, existe una matriz M^{-1} en $\mathrm{Mat}_2(A)$ tal que $MM^{-1} = I$, luego $|M||M^{-1}| = |I| = 1$, lo que prueba que $|M|$ es una unidad de A. ∎

Observemos ahora que si dos matrices cumplen $MN = I$, al tomar determinantes resulta que ambos son unidades de A, luego existe M^{-1} y

$$N = IN = M^{-1}MN = M^{-1}I = M^{-1}.$$

Así pues, para que una matriz sea la inversa de otra basta con que cumpla $MM^{-1} = I$, y automáticamente se cumplirá $M^{-1}M = I$ (y viceversa).

En estos términos, lo que sucede con la matriz (11.11) es que su determinante es $|M| = 2$, que es una unidad de $\mathbb{Q}$, pero no de $\mathbb{Z}$. Por eso es regular en $\mathrm{Mat}_2(\mathbb{Q})$, pero no en $\mathrm{Mat}_2(\mathbb{Z})$. En general, para que una matriz con coeficientes en un cuerpo tenga inversa, la condición necesaria y suficiente es que su determinante no se anule.

Definición 11.8 Si A es un dominio, llamamos $\mathrm{LG}_2(A)$ al conjunto de todas las matrices regulares con coeficientes en A, que es un grupo con el producto de matrices, llamado *grupo lineal general*. Similarmente se define el *grupo lineal especial* $\mathrm{LE}_2(A)$ formado por todas las matrices de determinante 1.

Es el primer ejemplo que encontramos de grupos no abelianos.

Conviene observar en qué se traduce concretamente que una matriz tenga determinante nulo:

Teorema 11.9 *Una matriz M con coeficientes en un dominio de factorización única A tiene determinante nulo si y sólo si es de la forma*

$$M = \begin{pmatrix} kr & ks \\ lr & ls \end{pmatrix},$$

para ciertos k, l, r, s en A tales que $(k, l) = 1$.

Demostración: Obviamente toda matriz de esta forma tiene determinante nulo. Supongamos ahora que

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

tiene determinante nulo, es decir, $ad = bc$. Si todas las entradas son nulas, la conclusión es obvia. En otro caso, por simetría podemos suponer que $c \neq 0$. Si $a = 0$, entonces $b = 0$ y la conclusión es trivial, sin más que tomar $k = 0$, $l = 1$. Supongamos, pues, que $a \neq 0$ y expresemos $a/c = k/l$, con $(k, l) = 1$. Entonces

$$\left(\frac{k}{l}c, \frac{k}{l}\,d\right) = \left(\frac{a}{c}c, \frac{a}{c}\,d\right) = (a, b).$$

Entonces $kc = la$, luego $l \mid kc$, luego $l \mid c$, e igualmente $l \mid d$. Pongamos que $c = lr$ y $d = ls$. La relación anterior nos da que $a = kr$ y $b = ks$. ∎

Nota El teorema anterior vale cuando A es un cuerpo. Esto puede verse como consecuencia de que todo cuerpo es trivialmente un dominio de factorización única (porque todo elemento no nulo es una unidad, luego no hay primos y no hay elementos que factorizar), pero también basta tener en cuenta la relación

$$\left(\frac{a}{c}c, \frac{a}{c}\,d\right) = (a, b),$$

que permite tomar $k = a/c$, $l = 1$, $r = c$, $s = d$. ∎

Sistemas de ecuaciones lineales La relación entre las matrices y los cambios de variables se puede hacer más estrecha si convenimos en identificar

$$(x, y) = \begin{pmatrix} x & y \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x & 0 \\ y & 0 \end{pmatrix}.$$

Así:

$$(x, y)\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ax + cy & bx + dy \\ 0 & 0 \end{pmatrix} = (ax + cy, bx + dy),$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by & 0 \\ cx+dy & 0 \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}.$$

Con estas definiciones es inmediato que el producto sigue siendo asociativo cuando el factor inicial es una matriz fila o el final una matriz columna, pues realmente se trata de productos de matrices cuadradas, y el producto de matrices es asociativo.

En estos términos, un sistema de ecuaciones lineales

$$\left.\begin{aligned} ax+cy &= u \\ bx+dy &= v \end{aligned}\right\}$$

con coeficientes en un dominio íntegro A puede expresarse como una ecuación matricial:

$$(x,y)M = (u,v)$$

donde M es la matriz de coeficientes y, si $|M| \neq 0$, entonces M tiene inversa en el cuerpo de cocientes de A, y la igualdad anterior equivale a

$$(x,y) = (u,v)M^{-1},$$

con lo que el sistema tiene solución única (aunque no necesariamente con valores en A). Si $|M| = 0$ y A es un dominio de factorización única (en particular, si es un cuerpo), el teorema 11.9 nos permite escribir el sistema como

$$\left.\begin{aligned} k(rx+sy) &= u \\ l(rx+sy) &= v \end{aligned}\right\}$$

Descartando los casos triviales, como $r = s = 0$ o $k = 0$, o $l = 0$, para que tenga soluciones en A es necesario que $k \mid u$, $l \mid v$ y $u/k = v/l$, lo que significa que las dos ecuaciones son la misma, y todo se reduce a encontrar las soluciones de una única ecuación lineal. En el caso en que $A = \mathbb{Z}$ hemos visto cómo hacerlo en la sección 2.2.

Ejemplo 1 Vamos a encontrar las soluciones enteras de la ecuación

$$2x^2 + 17xy + 35y^2 = 2.$$

Para ello observamos que el discriminante del miembro izquierdo es

$$D = 17^2 - 4\cdot 2\cdot 35 = 9,$$

que es un cuadrado perfecto, luego la forma cuadrática es reducible. Para descomponerla calculamos las raíces del polinomio $2x^2 + 17x + 35$, que son -5 y $-14/4$. Por lo tanto,

$$2x^2 + 17x + 35 = 2(x+5)(x+7/2) = (x+5)(2x+7),$$

y cambiando x por x/y pasamos a

$$2x^2 + 17xy + 35y^2 = (x+5y)(2x+7y).$$

Para resolver

$$(x+5y)(2x+7y)=2$$

consideramos todas las factorizaciones de $2=2\cdot1=1\cdot2=(-2)(-1)=(-1)(2)$, que nos llevan a cuatro sistemas de ecuaciones:

$$\left.\begin{matrix}x+5y=2\\2x+7y=1\end{matrix}\right\},\quad \left.\begin{matrix}x+5y=1\\2x+7y=2\end{matrix}\right\},\quad \left.\begin{matrix}x+5y=-2\\2x+7y=-1\end{matrix}\right\},\quad \left.\begin{matrix}x+5y=-1\\2x+7y=-2\end{matrix}\right\}$$

La matriz de coeficientes y su inversa son

$$M=\begin{pmatrix}1&2\\5&7\end{pmatrix},\qquad M^{-1}=-\frac{1}{3}\begin{pmatrix}7&-2\\-5&1\end{pmatrix},$$

luego las soluciones de los cuatro sistemas son

$$(x,y)=-\frac{1}{3}(2,1)\begin{pmatrix}7&-2\\-5&1\end{pmatrix}=(-3,1)$$

e igualmente $(x,y)=(1,0),(3,-1),(-1,0)$. Las cuatro soluciones han resultado ser enteras. Si alguna no lo hubiera sido, la habríamos descartado. ∎

Del mismo modo se puede resolver cualquier otra ecuación diofántica determinada por una forma cuadrática reducible en la que los sistemas de ecuaciones resultantes tengan matriz de coeficientes con determinante no nulo.

Ejemplo 2 Vamos a encontrar las soluciones enteras de la ecuación

$$9x^2-30xy+25y^2=4.$$

Su discriminante es $D=30^2-4\cdot9\cdot25=0$, luego es reducible. Para descomponerla calculamos la raíz del polinomio $9x^2-30x+25$, que es $5/3$, luego

$$9x^2-30x+25=9(x-5/3)^2=(3x-5)^2,$$

de donde[2]

$$9x^2-30xy+25y^2=(3x-5y)^2.$$

Por lo tanto, para que la ecuación tenga solución entera, es necesario que el término independiente sea un cuadrado, como lo es en este caso, y las soluciones son las mismas que las de las ecuaciones

$$3x-5y=\pm2.$$

Ejemplos de soluciones particulares son $(4,2)$ y $(1,1)$, y las soluciones de la ecuación $3x-5y=0$ son las de la forma $(5k,3k)$, luego las soluciones de la ecuación dada son

$$(x,y)=(4+5k,2+3k),(1+5k,1+3k).$$

∎

Por lo tanto, sólo queda por resolver el modo de encontrar las soluciones enteras de las ecuaciones definidas por formas cuadráticas irreducibles.

[2]Notemos que no es casual que factorice como un polinomio de grado 1 al cuadrado. Por el teorema 11.9 la factorización tiene que ser de la forma $(krx+ksy)(lrx+lsy)$, pero si la forma de partida es primitiva, necesariamente $k=l=\pm1$, luego la descomposición tiene que ser de la forma $\pm(rx+sy)^2$.

Expresión matricial de una forma cuadrática Observemos ahora que una forma cuadrática

$$ax^2 + bxy + cy^2$$

admite también una expresión matricial, a saber,

$$ax^2 + bxy + cy^2 = (x, y) \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

En efecto,

$$(x, y) \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (ax + \frac{b}{2}y, \frac{b}{2}x + cy) \begin{pmatrix} x \\ y \end{pmatrix}$$

$$= ax^2 + \frac{b}{2}xy + \frac{b}{2}xy + cy^2 = ax^2 + bxy + cy^2,$$

donde estamos adoptando el convenio de que

$$(a, b) \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c & 0 \\ d & 0 \end{pmatrix} = \begin{pmatrix} ac + bc & 0 \\ 0 & 0 \end{pmatrix} = ac + bd.$$

Así, cada forma cuadrática $ax^2 + bxy + cy^2$ está determinada por la matriz

$$M = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix},$$

cuyo determinante es $\Delta = ac - b^2/4$, que guarda una relación muy simple con el discriminante $D = 4\Delta$.

Notemos que cada matriz de esta forma determina una forma cuadrática, y que matrices distintas se corresponden con formas distintas, pues, si $f(x, y)$ es una forma cuadrática, sus coeficientes están determinados por

$$a = f(1, 0), \quad c = f(0, 1), \quad b = f(1, 1) - f(1, 0) - f(0, 1).$$

En estos términos es muy fácil determinar cómo se transforma una forma cuadrática cuando se le aplica un cambio de variables. Para ello conviene definir el concepto de matriz traspuesta:

Definición 11.10 La *matriz traspuesta* de una matriz

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

es la matriz

$$M^t = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

que resulta de cambiar sus filas por sus columnas.

Esta definición se aplica en particular a matrices fila y columna:

$$(a,b)^t = \begin{pmatrix} a \\ b \end{pmatrix}, \qquad \begin{pmatrix} a \\ b \end{pmatrix}^t = (a,b).$$

Una comprobación rutinaria muestra que

$$(MN)^t = N^t M^t.$$

También es obvio que $|M| = |M^t|$.

Ahora consideremos una forma cuadrática

$$f(x,y) = (x,y)M \begin{pmatrix} x \\ y \end{pmatrix}$$

y vamos a aplicarle un cambio de variables

$$(x,y) = (x',y')N,$$

que equivale a

$$\begin{pmatrix} x \\ y \end{pmatrix} = N^t \begin{pmatrix} x' \\ y' \end{pmatrix}$$

Al aplicar el cambio de variables obtenemos la forma cuadrática

$$f(x',y') = (x',y')NMN^t \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

Así pues:

Teorema 11.11 *Cuando a una forma cuadrática de matriz M se le aplica un cambio de variable de matriz N, la forma cuadrática resultante es la que tiene matriz NMN^t. En particular la nueva forma cuadrática tiene discriminante $D' = D|N|^2$, donde D es el discriminante de la forma dada.*

La última afirmación es consecuencia de que

$$|NMN^t| = |N||M||N^t| = |M||N|^2,$$

y basta multiplicar por 4 para tener la relación entre los discriminantes.

Ejemplo Ahora es fácil obtener la forma cuadrática que resulta de aplicarle a la forma (11.1) el primer cambio de variables de (11.3). Su matriz será

$$\begin{pmatrix} -1 & 3 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 13 & 5 \\ 5 & 2 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 3 & -2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 3 & -2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

por lo que la forma resultante es

$$(x',y') \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = x'^2 + y'^2.$$

■

11.3 Fracciones continuas finalmente iguales

Antes de aplicar los resultados sobre matrices al estudio de las formas cuadráticas, para no interrumpir la exposición posterior veremos aquí una aplicación a las fracciones continuas que vamos a necesitar. Se trata de determinar cuándo dos números irracionales α y β tienen fracciones continuas finalmente iguales, es decir:

$$\alpha = [a_0, \ldots, a_m, \gamma], \qquad \beta = [b_0, \ldots, b_n, \gamma].$$

El teorema 10.10 nos da que

$$\alpha = \frac{p_m\gamma + p_{m-1}}{q_m\gamma + q_{m-1}}, \qquad \beta = \frac{p'_n\gamma + p'_{n-1}}{q'_n\gamma + q'_{n-1}},$$

donde, según 10.3, se cumple que

$$p_m q_{m-1} - p_{m-1}q_m = \pm 1, \qquad p'_n q'_{n-1} - p'_{n-1}q'_n = \pm 1.$$

Despejando γ en la segunda igualdad, sustituyendo la expresión en la primera y operando llegamos a que

$$\alpha = \frac{a\beta + b}{c\gamma + d},$$

para ciertos números enteros que cumplen $ad - bc = \pm 1$. No obstante, como tendríamos que repetir cálculos similares muchas veces, conviene demostrar algunos resultados generales.

Definición 11.12 Si $\mathbb{I}$ es el conjunto de los números irracionales, una aplicación $f : \mathbb{I} \longrightarrow \mathbb{I}$ es una *transformación modular* si existen enteros a, b, c, d tales que $ad - bc = \pm 1$ de modo que, para todo irracional α,

$$f(\alpha) = \frac{a\alpha + b}{c\alpha + d}.$$

Notemos que no puede ser $c\alpha + d = 0$, pues entonces α sería racional.

Ahora observamos que la composición de dos transformaciones modulares es una transformación modular. Explícitamente, si

$$\beta = \frac{a\alpha + b}{c\alpha + d} \qquad \text{y} \qquad \gamma = \frac{a'\beta + b'}{c'\beta + d'},$$

entonces

$$\gamma = \frac{a'\frac{a\alpha+b}{c\alpha+d} + b'}{c'\frac{a\alpha+b}{c\alpha+d} + d'} = \frac{a'(a\alpha + b) + b'(c\alpha + d)}{c'(a\alpha + b) + d'(c\alpha + d)} = \frac{(a'a + b'c)\alpha + (a'b + b'd)}{(c'a + d'c)\alpha + (d'c + d'd)}.$$

Si comparamos este cálculo con

$$\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a'a + b'c & a'b + b'd \\ c'a + d'c & c'b + d'd \end{pmatrix},$$

vemos que al componer la transformación modular cuyos coeficientes forman una matriz M con otra cuyos coeficientes forman una matriz M', el resultado es la transformación modular cuyos coeficientes forman la matriz $M'M$. Notemos que la condición sobre los coeficientes que exige la definición es que $|M| = \pm 1$, $|M'| = \pm 1$, luego también se cumple que $|M'M| = \pm 1$, por lo que, ciertamente, la composición es modular.

La relación con las matrices tiene más consecuencias. Si una transformación modular f tiene matriz M, entonces, la matriz M^{-1} tiene también coeficientes enteros, y define otra transformación modular f^{-1} de modo que las composiciones $f \circ f^{-1}$ y $f^{-1} \circ f$ vienen definidas por la matriz I, luego ambas son la transformación modular $g(\alpha) = \alpha$. En otras palabras: toda transformación modular tiene una transformación inversa, que también es modular.

Esto lleva a la definición siguiente:

Definición 11.13 Dos números irracionales α y β son *equivalentes* si existen enteros a, b, c, d tales que

$$\alpha = \frac{a\beta + b}{c\beta + d}, \qquad ad - bc = \pm 1. \tag{11.12}$$

En otras palabras, dos números irracionales son equivalentes si y sólo si uno puede transformarse en el otro mediante una transformación modular. Es inmediato que la equivalencia de números irracionales tiene las propiedades reflexiva, simétrica y transitiva.

Los teoremas 10.3 y 10.10 nos dan que la transformación $\alpha = [a_0, \ldots, a_n, \beta]$ es modular, dada concretamente por

$$\alpha = \frac{\beta p_n + p_{n-1}}{\beta q_n + q_{n-1}}.$$

El teorema siguiente caracteriza las transformaciones modulares que se pueden expresar de esta forma:

Teorema 11.14 *Si una transformación modular* (11.12) *cumple* $c > d > 0$ *entonces se puede expresar de la forma* $\alpha = [a_0, \ldots, a_n, \beta]$ *para ciertos enteros* $a_0, \ldots, a_n$*, todos positivos salvo quizá el primero.*

Demostración: Hay que probar que existen $a_0, \ldots, a_n$ tales que

$$p_n = a, \quad p_{n-1} = b, \quad q_n = c, \quad q_{n-1} = d. \tag{11.13}$$

Lo probaremos por inducción sobre d.

Si $d = 1$ tenemos que $a = bc \pm 1$. En el caso $a = bc + 1$ sirve $\alpha = [b, c, \beta]$. Si se cumple $a = bc - 1$, entonces $\alpha = [b - 1, 1, c - 1, \beta]$.

Supongamos ahora que $d > 1$. Aplicando el teorema 10.2, las ecuaciones (11.13) equivalen a

$$p_{n-1} = b, \quad p_{n-2} = a - a_n b, \quad q_{n-1} = d, \quad q_{n-2} = c - a_n d. \tag{11.14}$$

Se sigue cumpliendo $b(c - a_n d) - (a - a_n b)d = \pm 1$ para cualquier a_n, y por hipótesis de inducción (11.14) tendrá solución si garantizamos $d > c - a_n d > 0$, o equivalentemente, si $c/d > a_n > (c-d)/d$.

Notemos que c/d no puede ser entero, pues si $c = kd$ entonces $d \mid 1$. Como $c/d - (c-d)/d = 1$, podemos tomar un número natural a_n en estas condiciones y así se cumple el teorema. ∎

Teorema 11.15 *Dos números irracionales α y β son equivalentes si y sólo si sus desarrollos en fracción continua son finalmente iguales, es decir, si*

$$\alpha = [a_0, \dots, a_m, c_0, c_1, \dots\,], \qquad \beta = [b_0, \dots, b_n, c_0, c_1, \dots\,].$$

Demostración: El teorema 10.10 nos da que en estas condiciones tanto α como β son equivalentes al número $[c_0, c_1, \dots\,]$, luego son equivalentes entre sí. Supongamos ahora que α y β son equivalentes. Digamos que

$$\alpha = \frac{a\beta + b}{c\beta + d}, \qquad ad - bc = \pm 1.$$

Podemos suponer que $c\beta + d > 0$. Sea $\beta = [b_0, \dots, b_{k-1}, \beta_k]$, donde $\beta_k = [b_k, b_{k+1}, \dots\,]$. Entonces:

$$\beta = \frac{\beta'_k p_{k-1} + p_{k-2}}{\beta'_k q_{k-1} + q_{k-2}}.$$

Componiendo las transformaciones modulares obtenemos que

$$\alpha = \frac{P\beta'_k + R}{Q\beta'_k + S},$$

donde

$$\begin{aligned} P &= ap_{k-1} + bq_{k-1}, \\ R &= ap_{k-2} + bq_{k-2}, \\ Q &= cp_{k-1} + dq_{k-1}, \\ S &= cp_{k-2} + dq_{k-2}, \end{aligned}$$

que son enteros y cumplen $PS - QR = \pm 1$.

Por 10.3, y puesto que β se encuentra entre dos convergentes consecutivos cualesquiera, $|p_{k-1}/q_{k-1} - \beta| < 1/q_{k-1}q_k$, o sea, $|p_{k-1} - \beta q_{k-1}| < 1/q_k$. Por lo tanto $p_{k-1} = \beta q_{k-1} + \delta/q_{k-1}$, e igualmente $p_{k-2} = \beta q_{k-2} + \delta'/q_{k-2}$, con $|\delta|, |\delta'| < 1$.

De aquí resulta que

$$Q = (c\beta + d)q_{k-1} + \frac{c\delta}{q_{k-1}}, \qquad S = (c\beta + d)q_{k-2} + \frac{c\delta'}{q_{k-2}}.$$

Teniendo en cuenta que $c\beta + d > 0$, es claro que haciendo k suficientemente grande podemos conseguir $Q > S > 0$. Aplicando el teorema anterior resulta que $\alpha = [a_0, \dots, a_m, \beta_k]$, de donde se sigue el teorema. ∎

11.4 Equivalencia de formas cuadráticas

Pasamos ya a abordar lo que es el objeto central de este capítulo. Partimos de la definición siguiente:

Definición 11.16 Dos formas cuadráticas con coeficientes enteros son *(estrictamente) equivalentes* si una puede transformarse en la otra mediante un cambio de variables de determinante ± 1 (de discriminante 1).

Así, si dos formas son estrictamente equivalentes, en particular son equivalentes. Luego veremos ejemplos de que hay formas equivalentes que no son estrictamente equivalentes, pero hay que tener presente que puede darse el caso que ilustra este ejemplo:

Ejemplo En las secciones precedentes hemos visto que la forma cuadrática

$$13x^2 + 10xy + 2y^2$$

es equivalente a $x^2 + y^2$, pues una puede transformarse en la otra mediante los cambios de variables (11.3), que tienen determinante -1, así que en principio no podemos decir que la equivalencia sea estricta. Ahora bien, el cambio de variables $y = x'$, $x = y'$ tiene matriz

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

luego su determinante es -1, y transforma la forma cuadrática $x^2 + y^2$ en ella misma, así que si componemos este cambio con el de (11.3), es decir, si consideramos el cambio

$$x = x' - y', \qquad y = -2x' + 3y',$$

cuya matriz es

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -1 & 3 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix},$$

tenemos un cambio de determinante 1 que produce el mismo efecto en la forma dada:

$$\begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} 13 & 5 \\ 5 & 2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Vemos así que el hecho de que un cambio de variable de determinante -1 transforme una forma cuadrática en otra no impide que pueda haber otro de determinante 1 que tenga el mismo efecto, por lo que las formas pueden ser también estrictamente equivalentes. ∎

He aquí algunas propiedades elementales de la equivalencia (estricta) de formas cuadráticas:

1. **Propiedad reflexiva** *Toda forma cuadrática es estrictamente equivalente a sí misma.*

   En efecto, una forma cuadrática se transforma en sí misma mediante el cambio de variables trivial, de matriz I, que tiene determinante 1.

2. **Propiedad simétrica** *Si una forma cuadrática f es (estrictamente) equivalente a otra g, entonces g es (estrictamente) equivalente a f.*

   Esto se debe a que los cambios de variable de determinante ± 1 tienen un cambio inverso con el mismo determinante, por lo que si un cambio transforma f en g, el cambio inverso transforma g en f.

3. **Propiedad transitiva** *Si una forma cuadrática f es (estrictamente) equivalente a otra g y g es (estrictamente) equivalente a h, entonces f es (estrictamente equivalente) a h.*

   Esto se debe a que si un cambio de variables transforma f en g y otro g en h, su composición transforma f en h, y el determinante de la composición es el producto de los determinantes.

4. *Dos formas cuadráticas equivalentes tienen el mismo discriminante. En particular, una es irreducible si y sólo si lo es la otra.*

   Esto se debe a que, en general, al aplicar un cambio de variables de matriz N a una forma cuadrática de discriminante D, obtenemos una forma cuadrática de discriminante $D|N|^2$, luego si $|N| = \pm 1$, el nuevo discriminante es también D.

5. *Dos formas cuadráticas equivalentes toman los mismos valores enteros.*

   Si f tiene matriz M y el cambio de variables tiene matriz N, tenemos que la matriz de g es NMN^t. Así, si $f(a,b) = c$, esto significa que

   $$(a,b)M\begin{pmatrix}a\\b\end{pmatrix} = c,$$

   pero llamando $(u,v) = (a,b)N^{-1}$, tenemos que $(a,b) = (u,v)N$, luego

   $$g(u,v) = (u,v)NMN^t\begin{pmatrix}u\\v\end{pmatrix} = (a,b)M\begin{pmatrix}a\\b\end{pmatrix} = c.$$

   Por lo tanto, g también toma el valor c. Intercambiando los papeles de f y g obtenemos el recíproco.

6. *Si dos formas cuadráticas son equivalentes y una es definida positiva, definida negativa o indefinida, lo mismo vale para la otra.*

   Esto es consecuencia inmediata de la propiedad anterior.

7. *Si dos formas cuadráticas son equivalentes y una es primitiva, la otra también lo es.*

Si $f(x,y) = ax^2 + bxy + cy^2$ es primitiva, entonces $f(1,0) = a$, $f(0,1) = c$ y $f(1,1) = a+b+c$ son tres valores que toma f y que no tienen divisores comunes. Si g es equivalente a f, entonces g toma esos tres mismos valores y, si g no fuera primitiva, es decir, si sus tres coeficientes fueran múltiplos de $d > 1$, entonces g sólo tomaría valores múltiplos de d, y no es el caso, luego g es primitiva.

Las tres primeras propiedades nos permiten hablar de clases de equivalencia y clases de equivalencia estricta de formas cuadráticas, exactamente igual que hablamos de clases de restos. La clase de equivalencia $[f]$ (estricta) de una forma cuadrática f contiene todas las formas cuadráticas (estrictamente) equivalentes a f, de modo que $[f] = [g]$ es otra forma de decir que f y g son (estrictamente) equivalentes.

Ahora bien, podemos definir una relación más elemental: diremos que dos formas cuadráticas *son del mismo orden* si tienen el mismo discriminante. De este modo, el conjunto de todas las formas cuadráticas con coeficientes enteros se divide en infinitos órdenes de formas cuadráticas, según su discriminante, y a su vez, cada orden de formas cuadráticas se divide en clases de equivalencia, y cada clase de equivalencia puede a su vez dividirse en dos clases de equivalencia estricta.

En efecto, si todas las formas de una clase de equivalencia $[f]$ son estrictamente equivalentes a f, entonces hay una única clase de equivalencia estricta, pero si en $[f]$ hay una forma g que es equivalente, pero no estrictamente equivalente a f, entonces $[f]$ se divide exactamente en dos clases de equivalencia estricta: la de las formas estrictamente equivalentes a f y la de las formas estrictamente equivalentes a g.

Esto es así porque, si una forma h de la clase $[f]$ no es estrictamente equivalente a f, es que el cambio de variable que la transforma en f tiene determinante -1, luego al componerlo con el cambio que transforma f en g (que también tiene determinante -1) obtenemos un cambio de determinante 1 que transforma h en g, luego h es estrictamente equivalente a g.

Esto plantea el problema de decidir, dadas dos formas cuadráticas, si son equivalentes o no, y, en caso afirmativo, de encontrar un cambio de variables que transforme una en otra. Para ello tenemos que tratar por separado el caso de las formas de discriminante negativo y el de las formas de discriminante positivo.

11.4.1 Formas definidas positivas

Ya hemos indicado que, al considerar formas cuadráticas $ax^2 + bxy + cy^2$ de discriminante negativo, no perdemos generalidad si las suponemos primitivas y definidas positivas, lo cual equivale a que $a, c > 0$ y, naturalmente, $D = b^2 - 4ac < 0$.

Ahora observamos que el cambio de variables

$$x = y', \qquad y = -x'$$

intercambia los coeficientes a y c y cambia b por $-b$, luego nos permite pasar a una forma equivalente en la que $a \leq c$. Por otra parte, el cambio

$$x = x' + ny', \qquad y = y'$$

la convierte en

$$ax^2 + (b + 2na)xy + (c + nb + n^2a)y^2,$$

con lo que, eligiendo n, podemos pasar a una forma estrictamente equivalente en la que $|b| \leq a$.

Con ello podemos perder la condición $a \leq c$, pero podemos repetir el proceso nuevamente, y tras un número finito de pasos (puesto que cada vez el valor de a se hace menor) llegamos a una forma estrictamente equivalente a la primera que cumple simultáneamente $|b| \leq a \leq c$. Más aún, si $b = -a$, el segundo cambio nos permite hacer $b = a$ sin cambiar c, y si $a = c$ entonces el primer cambio nos permite obtener $b \geq 0$.

La definición y el teorema siguientes recogen lo que hemos obtenido:

Definición 11.17 Una forma cuadrática definida positiva $ax^2 + bxy + cy^2$ está *reducida* si es primitiva y cumple $-a < b \leq a < c$ o bien $0 \leq b \leq a = c$.

Teorema 11.18 *Toda forma cuadrática primitiva definida positiva es estrictamente equivalente a una forma reducida.*

Más aún, tenemos un algoritmo para encontrar dicha forma.

Ejemplo Consideremos de nuevo la forma cuadrática

$$13x^2 + 10xy + 2y^2$$

que tiene discriminante $D = -4$. El algoritmo que hemos descrito da lugar a la sucesión siguiente:

$13x^2 + 10xy + 2y^2$	$x = y'$	$y = -x'$
$2x^2 - 10xy + 13y^2$	$x = x' + 3y'$	$y = y'$
$2x^2 + 2xy + y^2$	$x = y'$	$y = -x'$
$x^2 - 2xy + 2y^2$	$x = x' + y'$	$y = y'$
$x^2 + y^2$		

El cambio de variables correspondiente es

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 3 \\ -1 & 2 \end{pmatrix}$$

que se corresponde con las ecuaciones

$$x = -x' - y', \quad y = 3x' + 2y'.$$

Obtenemos así un nuevo cambio de variables, ligeramente distinto de los que habíamos considerado hasta ahora, que transforman la forma dada en $x^2 + y^2$. La diferencia es que éste lo hemos encontrado y en lugar de "sacárnoslo de la manga" oportunamente como habíamos hecho con los cambios anteriores. ∎

En el ejemplo anterior, si no nos hubiera interesado el cambio de variables, podríamos habernos ahorrado todo el proceso, porque x^2+y^2 es la única forma cuadrática reducida de discriminante -4! Vamos a probar algo más general:

Teorema 11.19 *Para cada discriminante $D<0$, existe un número finito de clases de equivalencia estricta de formas cuadráticas primitivas definidas positivas de discriminante D.*

DEMOSTRACIÓN: Basta observar que si $ax^2+bxy+cy^2$ es una forma cuadrática reducida de discriminante D, entonces $|b|\le a\le c$, luego

$$-D=4ac-b^2\ge 4a^2-a^2=3a^2,$$

luego $|b|\le a\le\sqrt{-D/3}$, lo que sólo deja un número finito de posibilidades para a y b, y la relación $D=b^2-4ac$ implica que, para cada una de ellas, hay a lo sumo una posibilidad para c. ∎

Conviene destacar de la demostración que toda forma cuadrática reducida de discriminante negativo debe cumplir

$$|b|\le a\le\sqrt{-D/3}, \tag{11.15}$$

pues esto nos permite encontrarlas todas. Por ejemplo, en el caso $D=-4$ una forma reducida debe cumplir $|b|\le a\le 1$, luego tiene que ser $a=1$, y entonces $-4=b^2-4c$ no puede cumplirse con $b^2=1$, luego tiene que ser $b=0$, y entonces $c=1$, luego la forma es x^2+y^2.

Concluimos que todas las formas cuadráticas (primitivas definidas positivas) de discriminante -4 son estrictamente equivalentes entre sí, luego los valores que toma cualquiera de ellas son los números naturales que son suma de dos cuadrados.

Ejemplo Vamos a calcular las formas cuadráticas (primitivas, definidas positivas) reducidas de discriminante $D=-20$.

Según (11.15), tienen que cumplir $|b|\le a\le 2$ y, por supuesto, $b^2-4ac=-20$. Si $a=1$ la ecuación $b^2-4c=-20$ no tiene solución si $b^2=1$, luego tiene que ser $b=0$, y entonces $c=5$, con lo que queda x^2+5y^2.

Si $a=2$ la ecuación $b^2-8c=-20$ no tiene solución si $b^2=0,1$, pero para $b^2=4$ vale $c=3$. Como $a<c$ tiene que ser $b\ge 0$, es decir, $b=2$.

Concluimos que hay sólo dos formas cuadráticas reducidas de discriminante $D=-20$, a saber,

$$x^2+5y^2,\qquad 2x^2+2xy+3y^2.$$

No son equivalentes entre sí, pues la segunda toma el valor 3, mientras que es claro que la primera no puede tomarlo.

Así pues, el orden de las formas cuadráticas de discriminante -20 se divide en dos clases de equivalencia de formas, cada una de las cuales consta de una única clase de equivalencia estricta.

Vemos, pues, que hemos encontrado "una compañera" a la forma cuadrática x^2+5y^2, y esto tiene una interpretación profunda. En la página 309 estudiamos los números naturales de la forma x^2+5y^2 y vimos que "si $\mathbb{Q}\big(\sqrt{-5}\,\big)$ tuviera factorización única", deberían ser los números tales que los primos que los dividen con exponente impar cumplen

$$p = 2, 5 \qquad \text{o} \qquad p \equiv 1, 3, 7, 9 \text{ (mód 20)}.$$

Sin embargo, no es cierto que $\mathbb{Q}\big(\sqrt{-5}\,\big)$ tenga factorización única y la realidad es otra. Las tablas que mostramos invitaban a conjeturar que un número es de la forma x^2+5y^2 si y sólo si lo es su parte libre de cuadrados, y que un número libre de cuadrados de esta forma es producto de los primos indicados, pero no todos los casos son posibles. En particular, los primos de la forma x^2+5y^2 parecen ser únicamente los que cumplen

$$p = 5 \qquad \text{o} \qquad p \equiv 1, 9 \text{ (mód 20)}.$$

He aquí ahora la lista de los primeros números de la forma $2x^2+2xy+3y^2$:

2	**3**	**7**	8	10	12	15	18	**23**	27	28	32	35	40	42	**43**
47	48	50	58	60	63	**67**	72	75	82	**83**	87	90	92	98	**103**
107	108	112	115	122	123	**127**	128	135	138	140	147	160	162	**163**	**167**
168	172	175	178	183	188	192	200	202	203	207	210	215	218	**223**	**227**
232	235	240	242	243	250	252	258	**263**	267	268	282	**283**	287	288	290
298	300	303	**307**	315	322	327	328	332	335	338	343	**347**	348	360	362
363	**367**	368	375	378	**383**	387	392	402	410	412	415	423	427	428	432
435	**443**	447	448	450	458	460	**463**	**467**	482	483	**487**	488	490	492	498

Vemos que en ella aparecen el 2 y los primos $p \equiv 3, 7$ (mód 20), que son precisamente los que echábamos en falta al estudiar los números de la forma x^2+5y^2. Si nos quedamos con la parte libre de cuadrados obtenemos:

2	**3**	**7**	2	**10**	3	**15**	2	**23**	23	7	2	**35**	10	**42**	**43**
47	3	2	**58**	15	7	**67**	2	3	**82**	**83**	**87**	10	23	2	**103**
107	3	7	**115**	**122**	**123**	**127**	2	15	**138**	35	3	10	2	**163**	**167**
42	43	7	**178**	**183**	47	3	2	**202**	**203**	23	**210**	**215**	**218**	**223**	**227**
58	**235**	15	2	3	10	7	**258**	**263**	**267**	67	**282**	**283**	**287**	2	**290**
298	3	**303**	**307**	35	**322**	**327**	82	83	**335**	2	7	**347**	87	10	**362**
3	**367**	23	15	42	**383**	43	2	**402**	**410**	103	**415**	47	**427**	107	3
435	**443**	**447**	7	2	**458**	115	**463**	**467**	**482**	**483**	**487**	122	10	123	**498**

De nuevo podemos conjeturar que un número es de la forma $2x^2+2xy+3y^2$ si y sólo si lo es su parte libre de cuadrados, y además podemos apreciar que cualquier número formado por primos $p = 2, 5$ o $p \equiv 1, 3, 7, 9$ (mód 20) aparece en una de las dos tablas. Todo esto se resume en la conjetura siguiente:

> *Un número natural está representado por una forma cuadrática de discriminante $D = -20$ (o, equivalentemente, es de una de las formas $n = x^2+5y^2$ o bien $n = 2x^2+2xy+3y^2$), si y sólo si los primos que lo dividen con exponente impar cumplen*
>
> $$p = 2, 5 \qquad o \qquad p \equiv 1, 3, 7, 9 \text{ (mód 20)}.$$

¿Puede el lector conjeturar además algo sobre cómo se distribuyen los números entre las dos formas? ¿De qué forma es el producto de dos números de la forma x^2+5y^2? ¿Y el de dos números de la forma $2x^2+2xy+3y^2$? ¿Y el de un número de cada forma?

El teorema siguiente demuestra una parte de la conjetura anterior. ∎

Teorema 11.20 *Un entero r es de la forma $r = f(x_0, y_0)$, donde f es una forma cuadrática de discriminante $D \neq 0$ y $(x_0, y_0) = 1$ si y sólo si D es un resto cuadrático módulo $4r$.*

En particular, si r es libre de cuadrados (y, más en particular, si es primo), tenemos que r es de la forma $r = f(x_0, y_0)$, donde f es una forma cuadrática de discriminante D si y sólo si D es un resto cuadrático módulo $4r$. En efecto, si r es libre de cuadrados la condición $(x_0, y_0) = 1$ es redundante, pues si $d = (x_0, y_0)$, es claro que $d^2 \mid r$, luego tiene que ser $d = 1$.

Si aplicamos este resultado al ejemplo anterior, vemos que un primo p es de la forma $p = x^2 + 5y^2$ o de la forma $p = 2x^2 + 2xy + 3y^2$ si y sólo si $p = 2, 5$ o bien

$$\left(\frac{-20}{p}\right) = \left(\frac{-5}{p}\right) = 1,$$

lo cual a su vez equivale a que

$$\left(\frac{p}{5}\right) = (-1)^{(p-1)/2},$$

y es fácil ver que esto equivale a que p sea de la forma indicada en la conjetura.

Ejercicio: Probar que un número libre de cuadrados es de la forma $n = x^2 + 5y^2$ o bien $n = 2x^2 + 2xy + 3y^2$ si y sólo si cada uno de sus factores primos es de una de estas formas.

Demostración: Si $r = f(x_0, y_0)$ con $(x_0, y_0) = 1$, existen enteros u y v tales que $ux_0 - vy_0 = 1$. Consideramos la forma $g(x, y) = f(x_0x + vy, y_0x + uy)$, que es equivalente a f, luego también tiene discriminante D. Pero además $g(1, 0) = r$, luego

$$g(x, y) = rx^2 + sxy + tu^2,$$

luego $D = s^2 - 4rt$, luego $D \equiv s^2 \pmod{4r}$.

Recíprocamente, si $D = s^2 + 4rt$, entonces $f(x, y) = rx^2 + sxy - ty^2$ es una forma cuadrática de discriminante D que cumple $f(1, 0) = r$. ∎

Ejemplo Consideremos ahora el caso de discriminante $D = -56$. Es fácil concluir que hay cuatro formas reducidas:

$$x^2 + 14y^2, \qquad 2x^2 + 7y^2, \qquad 3x^2 + 2xy + 5y^2, \qquad 3x^2 - 2xy + 5y^2.$$

Ejercicio: Comprobar que existen tres clases de equivalencia no estricta de formas de discriminante -56.

Si el lector hace una exploración similar a la que hemos hecho en el ejemplo anterior, descubrirá que ya no es cierto que un número natural está representado por una de estas formas si y sólo si lo está su parte libre de cuadrados. Por ejemplo:

$$2^2 + 14 \cdot 1^2 = 3^2 \cdot 2, \qquad 7^2 + 14 \cdot 1^2 = 3^2 \cdot 7, \qquad 17^2 + 14 \cdot 5^2 = 3^2 \cdot 71,$$

pero las partes libres de cuadrados correspondientes no son de la forma x^2+14y^2.

¿Puede el lector establecer una conjetura sobre qué números están representados por alguna de las cuatro formas anteriores? ∎

Ha sido fácil justificar que las formas $x^2 + 5y^2$ y $2x^2 + 2xy + 3y^2$ no son equivalentes, pero conviene observar que el hecho de que no sean estrictamente equivalentes es consecuencia de un hecho general:

Teorema 11.21 *Dos formas reducidas son estrictamente equivalentes si y sólo si son iguales.*

DEMOSTRACIÓN: Sea $f(x, y) = ax^2 + bxy + cy^2$ una forma reducida. Si $0 < |y| \leq |x|$ entonces

$$\begin{aligned} f(x,y) &\geq ax^2 - |bxy| + cy^2 = |x|(a|x| - |by|) + c|y|^2 \\ &\geq |x|^2(a - |b|) + c|y|^2 \geq a - |b| + c. \end{aligned}$$

Se obtiene el mismo resultado si suponemos $0 < |x| \leq |y|$.

Puesto que $a - |b| + c$ se alcanza en $(1, \pm 1)$, tenemos que esta cantidad es el mínimo de f sobre pares (x, y) donde $x \neq 0$, $y \neq 0$.

Si consideramos tan sólo pares (x, y) de enteros primos entre sí, los únicos que falta por considerar aparte de los que tienen componentes no nulas son $(1, 0)$ y $(0, 1)$, donde f toma los valores a y c. Por lo tanto, el conjunto de las imágenes que toma f sobre tales pares comienza con $a \leq c \leq a - |b| + c, \ldots$

Es fácil ver que un cambio de variables de determinante 1 biyecta los pares de números enteros primos entre sí, luego si dos formas cuadráticas reducidas de coeficientes (a, b, c) y (a', b', c') son estrictamente equivalentes, ambas alcanzan el mismo mínimo sobre tales pares, es decir, $a = a'$.

Si $a = b = c$, entonces la primera forma toma el valor a al menos sobre tres pares de enteros primos entre sí. Si $a = c \neq b$ (y entonces $c < a - |b| + c$) lo toma sólo dos veces y si $a < c$ lo toma sólo una vez. Lo mismo le ocurre a la segunda forma, luego si $a = b = c$ tenemos que $a' = b' = c'$ y ambas son la misma forma, si $a = c \neq b$ tenemos $a' = c' \neq b'$, y si $a < c$ entonces también $a' < c'$, y en este último caso c y c' son ambos iguales al mínimo valor distinto de a que toman ambas formas sobre pares de enteros primos entre sí. En cualquier $c = c'$.

Finalmente, si $a = c \neq b$ o bien $a < c$, concluimos por el mismo argumento que también $a - |b| + c = a' - |b'| + c'$, y así en cualquier caso $b = \pm b'$.

Vamos a ver que si $b = -b'$ entonces $b = 0$. En el caso $a = c$ es inmediato por la definición de forma reducida (sería, $b \geq 0$, $b' \geq 0$), luego suponemos $a < c$.

No puede ser $b = a$ porque entonces $b' = -a'$, en contra de la definición de forma reducida. Así pues, $-a < b < a < c$. Llamemos f a la primera

forma y f' a la segunda. Digamos que $f(x,y) = f'(px+qy, rx+sy)$. Entonces $a = f(1,0) = f'(p,r)$, pero f' sólo toma el valor a en $(\pm 1, 0)$, luego $p = \pm 1$ y $r = 0$. Como $ps - qr = 1$, ha de ser $s = \pm 1$. Igualmente $c = f(0,1) = f'(q,s)$, luego $q = 0$. En definitiva, $f(x,y) = f'(\pm x, \pm y)$, de donde $b = b' = 0$. ∎

Esto resuelve completamente el problema de determinar si dos formas cuadráticas de discriminante negativo son o no equivalentes, o estrictamente equivalentes. Para la equivalencia estricta basta reducir ambas formas, y serán estrictamente equivalentes si y sólo si sus reducciones son iguales. Para saber si son equivalentes (en caso de que no sean estrictamente equivalentes) basta estudiar si una es estrictamente equivalente a la forma que resulta de aplicar cualquier cambio de variables de discriminante negativo a la otra.

Ejercicio: Encontrar todas las formas cuadráticas reducidas de discriminante -23. Estudiar si son equivalentes.

11.4.2 Formas indefinidas

La equivalencia (estricta) de formas cuadráticas de discriminante positivo puede estudiarse en términos análogos a los que hemos empleado para tratar el caso de discriminante negativo, pero la situación es un poco más complicada. Para empezar, necesitamos una definición distinta de "forma reducida":

Definición 11.22 Una forma cuadrática irreducible primitiva $ax^2 + bxy + cy^2$ de discriminante $D > 0$ está *reducida* si

$$0 < b < \sqrt{D} \qquad \text{y} \qquad \sqrt{D} - b < 2|a| < \sqrt{D} + b.$$

Conviene observar que toda forma cuadrática reducida cumple las propiedades siguientes:

1. $ac < 0$.

   Esto se debe a que $0 < b < \sqrt{D}$, luego $ac = (b^2 - D)/4 < 0$.

2. $\sqrt{D} - b < 2|c| < \sqrt{D} + b$.

   Tenemos que

   $$(\sqrt{D} - b)(\sqrt{D} + b) = D - b^2 = -4ac = 2|a|2|c|,$$

   donde todos los factores son positivos. Como $\sqrt{D} - b < 2|a|$,

   $$2|a|2|c| = (\sqrt{D} - b)(\sqrt{D} + b) < 2|a|(\sqrt{D} + b),$$

   luego $2|c| < \sqrt{D} + b$. Similarmente, como $2|a| < \sqrt{D} + b$,

   $$2|a|(\sqrt{D} - b) < (\sqrt{D} - b)(\sqrt{D} + b) = 2|a|2|c|,$$

   luego $\sqrt{D} - b < 2|c|$.

3. $|a| < \sqrt{D}$, $|c| < \sqrt{D}$.
   Como $2|a| < \sqrt{D}+b$ y $0 < b < \sqrt{D}$, tenemos que $2|a| < 2\sqrt{D}$, e igualmente con c.

Como consecuencia de 2. y 3. sólo hay una cantidad finita de formas cuadráticas reducidas con un mismo discriminante.

Por otra parte:

Teorema 11.23 *Toda forma cuadrática irreducible primitiva de discriminante $D > 0$ es estrictamente equivalente a una forma reducida.*

DEMOSTRACIÓN: Sea $ax^2 + bxy + cy^2$ una forma en las condiciones del enunciado. Al aplicarle el cambio de variables

$$x = -y', \qquad y = x' + ny'$$

se transforma en

$$cx^2 + (-b + 2nc)xy + (a - nb + n^2c)y^2.$$

Por lo tanto, eligiendo n de modo que

$$\sqrt{D} - 2|c| < -b + 2nc < \sqrt{D},$$

pasamos a una forma estrictamente equivalente tal que

$$\sqrt{D} - 2|a| < b < \sqrt{D}.$$

Vamos a probar que si $|a| \leq |c|$, entonces la forma obtenida está reducida. En caso contrario, es decir, si $|c| < |a|$ podemos repetir el proceso. Como en cada paso se reduce el valor de $|a|$, tras un número finito de pasos tenemos que llegar a una forma reducida.

Tenemos que $0 < \sqrt{D} - b < 2|a|$. Como $D^2 - b = (\sqrt{D} - b)(\sqrt{D} + b)$ y un factor es positivo, el otro también lo es, es decir, $\sqrt{D} + b > 0$. Ahora:

$$2|c|(\sqrt{D} - b) < 4|a||c| = |D - b^2| = (\sqrt{D} - b|)(\sqrt{D} + b),$$

luego $2|c| \leq \sqrt{D} + b$ y en total

$$0 < \sqrt{D} - b < 2|a| \leq 2|c| \leq \sqrt{D} + b.$$

Esto implica en particular que $b > 0$ y así la forma cuadrática es reducida. ∎

Ejemplo Vamos a probar que las formas cuadráticas $77x^2 + 76xy + 17y^2$ y $2x^2 + 22xy - 7y^2$ son estrictamente equivalentes, y vamos a encontrar un cambio de variables que transforme la primera en la segunda.

Como la segunda ya está reducida, sólo tenemos que reducir la primera:

$77x^2+76xy+17y^2$	$x=-y' \quad y=x'+2y'$
$17x^2-8xy-7y^2$	$x=-y' \quad y=x'-y'$
$-7x^2+22xy+2y^2$	$x=-y' \quad y=x'+11y'$
$2x^2+22xy-7y^2$	

Para componer los cambios de variables multiplicamos las matrices:

$$\begin{pmatrix} 0 & 1 \\ -1 & 11 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}=\begin{pmatrix} 1 & -3 \\ 12 & -35 \end{pmatrix}$$

Por lo tanto, la primera forma cuadrática se transforma en la segunda mediante el cambio:

$$x=x'+12y', \qquad y=-3x'-35y'.$$

■

Ejercicio: Reducir la forma cuadrática x^2-10y^2 y encontrar un cambio de variables de determinante 1 que la transforme en una forma reducida equivalente.

Otra consecuencia inmediata de los hechos precedentes es la siguiente:

Teorema 11.24 *Para cada discriminante $D>0$, existe un número finito de clases de equivalencia estricta de formas cuadráticas irreducibles primitivas de discriminante D.*

Ejemplo Teniendo en cuenta las cotas que hemos obtenido tras la definición de forma reducida, es fácil encontrar todas las formas cuadráticas reducidas de discriminante $D=40$. Hay ocho en total:

$$\begin{array}{cccc} x^2+6xy-y^2, & -x^2+6xy+y^2, & 2x^2+4xy-3y^2, & -2x^2+4xy+3y^2, \\ 3x^2+2xy-3y^2, & -3x^2+2xy+3y^2, & 3x^2+4y^2-2y^2, & -3x^2+4y^2+2y^2. \end{array}$$

■

Sin embargo, muchas de las formas reducidas del ejemplo anterior son estrictamente equivalentes entre sí, por lo que no se cumple el análogo al teorema 11.21. Más precisamente, vamos a ver que si, al reducir una forma cuadrática, seguimos aplicando el algoritmo cuando ya hemos llegado a una forma cuadrática reducida, obtenemos más (al menos, siempre una más). Para ello conviene probar un resultado general:

Definición 11.25 Dos formas cuadráticas f y g son *adyacentes* (o, más precisamente, f es *adyacente por la izquierda* a g o g es *adyacente por la derecha* a f) si g se obtiene de f mediante un cambio de variables de la forma

$$x=-y', \qquad y=x'+ny',$$

donde n es un entero no nulo o, equivalentemente, f se obtiene de g mediante el cambio inverso

$$x=nx'+y', \qquad y=-x'.$$

Explícitamente, si $f(x,y) = ax^2 + bxy + cy^2$, entonces

$$g(x,y) = cx^2 + (-b+2nc)xy + (a - nb + n^2c)y^2.$$

o, equivalentemente, si $g(x,y) = a'x^2 + b'xy + c'y^2$,

$$f(x,y) = (c' - nb' + n^2a')x^2 + (-b' + 2na')xy + a'y^2.$$

En estos términos, la demostración del teorema 11.23 nos muestra cómo llegar a una forma cuadrática reducida estrictamente equivalente otra dada pasando sucesivamente a formas adyacentes por la derecha elegidas adecuadamente. Ahora vamos a ver que toda forma reducida tiene siempre una adyacentex reducida, tanto por la izquierda como por la derecha. Más aún:

Teorema 11.26 *Toda forma cuadrática reducida de discriminante negativo tiene una única adyacente por la derecha y una única adyacente por la izquierda reducida.*

DEMOSTRACIÓN: Vamos a probarlo para el caso de la derecha, pues el caso opuesto es totalmente análogo. Pongamos que

$$f(x,y) = ax^2 + bxy + cy^2$$

es una forma reducida y vamos a probar que existe un único valor de n no nulo tal que la forma

$$g(x,y) = cx^2 + (-b+2nc)xy + (a - nb + n^2c)y^2$$

es reducida. Esto equivale a que

$$0 < -b + 2nc < \sqrt{D}, \qquad \sqrt{D} + b - 2nc < 2|c| < \sqrt{D} - b + 2nc.$$

Las desigualdades segunda y tercera equivalen a

$$\sqrt{D} - 2|c| < -b + 2nc < \sqrt{D},$$

y existe un único n que cumple esta condición. Por lo tanto, sólo tenemos que demostrar que la forma obtenida con el valor de n determinado por estas desigualdades es reducida. De las cuatro desigualdades que tenemos que probar, la segunda y la tercera se cumplen por la elección de n.

En primer lugar observamos que, por la segunda relación demostrada antes del teorema 11.24 para los coeficientes de una forma reducida,

$$0 < (\sqrt{D} + b - 2|c|) + (2|c| - \sqrt{D} - b + 2nc) = 2nc.$$

En particular, esto prueba que $n \neq 0$. Para probar la primera desigualdad observamos que

$$0 < (\sqrt{D} - b) + (2|c| - \sqrt{D} - b + 2nc) + 2|c|(|n| - 1) = 2(-b + 2nc),$$

luego $-b + 2nc > 0$.

Por último:

$$0 < (\sqrt{D} - b) + 2|c|(|n| - 1) = \sqrt{D} - b + 2cn - 2|c|,$$

que es la cuarta desigualdad. ∎

Notemos que el criterio con el que elegimos n en la prueba del teorema anterior es el mismo considerado en la prueba del teorema 11.23, por lo que ambos teoremas se pueden combinar en el enunciado siguiente:

- *Si, a partir de una forma cuadrática, vamos pasando sucesivamente a formas adyacentes por la derecha eligiendo n con el criterio*

  $$\sqrt{D} - 2c < -b + 2nc < \sqrt{D},$$

  al cabo de un número finito de pasos llegamos a una forma reducida.

- *Si continuamos iterando, vamos obteniendo nuevas formas reducidas adyacentes.*

Ejemplo Si prolongamos la sucesión de formas que parte de $x^2 - 10y^2$ obtenemos

$x^2 - 10y^2$	$x = -y' \quad y = x'$
$-10x^2 + y^2$	$x = -y' \quad y = x' + 3y'$
$x^2 + 6xy - y^2$	$x = -y' \quad y = x' - 6y'$
$-x^2 + 6xy + y^2$	$x = -y' \quad y = x' + 6y'$
$x^2 + 6xy - y^2$	

y a partir de ahí entramos en un ciclo de formas de longitud 2. Similarmente, si partimos de la forma reducida $2x^2 + 4xy - 3y^2$ obtenemos:

$2x^2 + 4xy - 3y^2$	$x = -y' \quad y = x' - y'$
$-3x^2 + 2xy + 3y^2$	$x = -y' \quad y = x' + y'$
$3x^2 + 4xy - 2y^2$	$x = -y' \quad y = x' - 2y'$
$-2x^2 + 4xy + 3y^2$	$x = -y' \quad y = x' + y'$
$3x^2 + 2xy - 3y^2$	$x = -y' \quad y = x' - y'$
$-3x^2 + 4xy + 2y^2$	$x = -y' \quad y = x' + 2y'$
$2x^2 + 4xy - 3y^2$	

y entramos en un ciclo de longitud 6. En particular vemos que, a lo sumo, hay dos clases de equivalencia estricta de formas de discriminante -40, con representantes

$$x^2 + 6xy - y^2, \qquad 2x^2 + 4xy - 3y^2.$$

Ejercicio: Probar que estas dos formas no son equivalentes. AYUDA: Calcular los valores que toma la primera módulo 5. ∎

Ejercicio: Probar que la forma $191x^2 + 68xy + 6y^2$ es estrictamente equivalente a $x^2 + 6xy - y^2$ y encontrar un cambio de variables que transforme la primera en la segunda.

En general, puesto que sólo hay un número finito de formas cuadráticas reducidas de un mismo determinante, al calcular la sucesión de formas adyacentes reducidas, tarde o temprano se tiene que repetir una y, a partir de ese momento, la sucesión entra en un ciclo. Ahora bien, la forma repetida tiene que ser precisamente la primera de la sucesión (es decir, que el ciclo abarca todas las formas de la sucesión), pues en caso contrario, si la sucesión fuera, digamos,

$$f_1, \quad f_2, \quad f_3, \quad f_4, \quad f_5, \quad f_6, \quad f_7, \quad f_8, \quad f_3, \quad f_4, \quad \dots$$

donde todas las formas desde f_1 hasta f_8 son distintas dos a dos, tendríamos que f_2 y f_8 serían dos formas distintas adyacentes a f_3 por la izquierda, y hemos probado que esto es imposible.

Por consiguiente, las formas cuadráticas reducidas de un mismo discriminante se dividen en ciclos disjuntos de formas adyacentes. Para $D = -40$ hemos visto que hay dos ciclos: uno de longitud 2 y otro de longitud 6.

Otra observación elemental es que cada ciclo tiene longitud par, pues en cada paso el coeficiente a cambia de signo, luego hay que dar un número par de pasos para que una forma pueda repetirse. A su vez, esto implica que el número de clases de equivalencia estrictas de formas cuadráticas de un discriminante dado es siempre par.

Para terminar vamos a probar que dos formas cuadráticas reducidas de un mismo discriminante son estrictamente equivalentes si y sólo si están en el mismo ciclo de formas reducidas, con lo que tendremos un criterio sencillo para comprobar si dos formas cualesquiera son estrictamente equivalentes o no (basta iterarlas hasta llegar al ciclo de formas adyacentes reducidas y comprobar si llegamos o no al mismo ciclo).

Para ello necesitamos una caracterización de la equivalencia de formas cuadráticas:

Definición 11.27 Si $f(x,y) = ax^2 + bxy + cy^2$ es una forma cuadrática irreducible de discriminante $D > 0$, su *raíz asociada* es el número real

$$\alpha = \frac{-b - \sqrt{D}}{2a},$$

que es una de las raíces del polinomio $ax^2 + bx + c$.

Teorema 11.28 *Dos formas cuadráticas irreducibles de discriminante $D > 0$ son iguales si y sólo si tienen la misma raíz asociada.*

Demostración: Pongamos que

$$f(x,y) = ax^2 + bxy + cy^2, \qquad g(x,y) = a'x^2 + b'xy + c'y^2.$$

Si

$$\frac{-b - \sqrt{D}}{2a} = \frac{-b' - \sqrt{D}}{2a'},$$

tenemos un elemento del cuerpo $\mathbb{Q}\big(\sqrt{D}\,\big)$ y, por la unicidad de su representación en la forma $r + s\sqrt{D}$, tiene que ser

$$-\frac{b}{2a} = -\frac{b'}{2a'}, \qquad -\frac{1}{2a} = -\frac{1}{2a'},$$

luego $a = a'$, de donde $b = b'$ y, como $b^2 - 4ac = b'^2 - 4a'c'$, también $c = c'$. ∎

Más explícitamente, a partir de α podemos recuperar la forma que le corresponde teniendo en cuenta que

$$ax^2 + bx + c = a(x - \alpha)(x - \bar{\alpha}),$$

por lo que

$$ax^2 + bxy + cy^2 = a(x - y\alpha)(x - y\bar{\alpha}),$$

y $|a|$ se determina como el único que hace que la forma cuadrática obtenida tenga coeficientes enteros primos entre sí, mientras que el signo de a es el necesario para que la raíz asociada sea α y no $\bar{\alpha}$.

Ejemplo Si $f(x, y) = 2x^2 + 4xy - 3y^2$, entonces

$$\alpha = \frac{-4 - \sqrt{40}}{4} = \frac{-2 - \sqrt{10}}{2}$$

y, recíprocamente, a partir de α podemos calcular

$$(x - \alpha y)(x - \bar{\alpha} y) = x^2 + 2xy - \frac{3}{2}y^2,$$

luego tenemos que multiplicar por $a = 2$ para obtener una forma reducida, y así llegamos a la forma inicial. ∎

Teorema 11.29 *Si una forma cuadrática f se transforna en otra g mediante un cambio de variables*

$$(x, y) = (x', y') \begin{pmatrix} r & r' \\ s & s' \end{pmatrix}$$

de determinante 1*, entonces sus raíces asociadas son equivalentes, a través de la transformación modular*

$$\alpha_f = \frac{r\alpha_g + s}{r'\alpha_g + s'}.$$

Demostración: Pongamos que

$$f(x, y) = ax^2 + bxy + cy^2, \qquad g(x, y) = a'x^2 + b'xy + c'y^2.$$

El cambio de variables inverso es

$$(x', y') = (x, y) \begin{pmatrix} s' & -r' \\ -s & r \end{pmatrix}$$

de donde

$$f(x,y) = g(s'x - sy, -r'x + ry),$$

luego

$$\begin{aligned} a &= f(1,0) = g(s',-r') = a's'^2 - b'r's' + c'r'^2, \\ c &= f(0,1) = g(-s,r) = a's^2 - b'rs + c'r^2, \\ b &= f(1,1) - a - b = g(s'-s, r-r') - a - b \\ &= a'(s'-s)^2 + b'(s'-s)(r-r') + c'(r-r')^2 \\ &\quad -a's'^2 + b'r's' - c'r'^2 - a's^2 + b'rs - c'r^2 \\ &= -2a'ss' + b'(rs' + r's) - 2c'rr'. \end{aligned}$$

Por otra parte:

$$\begin{aligned} \frac{r\alpha_g + s}{r'\alpha_g + s'} &= \frac{r\frac{-b'-\sqrt{D}}{2a'} + s}{r'\frac{-b'-\sqrt{D}}{2a'} + s'} = \frac{-b'r + 2a's - r\sqrt{D}}{-b'r' + 2a's' - r'\sqrt{D}} \\ &= \frac{(-b'r + 2a's - r\sqrt{D})(-b'r' + 2a's' + r'\sqrt{D})}{(-b'r' + 2a's')^2 - r'^2(b'^2 - 4a'c')} \\ &= \frac{-b'(rs' + r's) + 2a'ss' + 2c'rr' - (rs' - r's)\sqrt{D}}{2(a's'^2 - b'r's' + c'r'^2)} \\ &= \frac{-b'(rs' + r's) + 2a'ss' + 2c'rr' - \sqrt{D}}{2(a's'^2 - b'r's' + c'r'^2)} = \frac{-b - \sqrt{D}}{2a}. \end{aligned}$$

■

Ahora observamos que si $ax^2 + bxy + cy^2$ es una forma cuadrática reducida y α es su raíz asociada, entonces, como $b > 0$, se cumple que

$$|\alpha| = \frac{b + \sqrt{D}}{2|a|}$$

y las condiciones de la definición de forma reducida implican que $|\alpha| > 1$, así como que $-1 < \overline{|\alpha|} < 0$, es decir, que $|\alpha|$ es un irracional cuadrático reducido en el sentido introducido tras el teorema 10.11. Dicho teorema implica que la fracción continua de $|\alpha|$ es periódica.

Ejemplo Si $f(x,y) = 2x^2 + 4xy - 3y^2$, tenemos que

$$\alpha = \frac{-2 - \sqrt{10}}{2},$$

luego

$$|\alpha| = \frac{2 + \sqrt{10}}{2} = [\overline{2,1,1}]$$

■

Sea $a'x^2 + b'xy + ay^2$ la forma cuadrática reducida adyacente a la anterior por la izquierda y sea α' su raíz asociada. Esto significa que el cambio de variables de matriz

$$\begin{pmatrix} 0 & 1 \\ -1 & n \end{pmatrix},$$

donde $n = (b + b')/(2a)$, transforma ésta forma en la precedente. El teorema anterior nos da que

$$\alpha' = \frac{-1}{\alpha + n}$$

o, equivalentemente,

$$\alpha = -n - \frac{1}{\alpha}\,'.$$

Si $a' > 0$, entonces $a < 0$, $n < 0$, $\alpha' < 0$ y $\alpha > 0$, luego

$$|\alpha| = |n| + \frac{1}{|\alpha'|}.$$

Si $a' < 0$ entonces $a > 0$, $n > 0$, $\alpha' > 0$ y $\alpha < 0$, luego llegamos a la misma relación entre los valores absolutos. Esto significa que

$$\alpha = [n, \alpha'],$$

o, en otras palabras, que si al desarrollo en fracción continua de $|\alpha|$ le quitamos su primer coeficiente, obtenemos el desarrollo en fracción continua de $|\alpha'|$.

Ejemplo Partiendo de que la forma cuadrática $2x^2+4xy-3y^2$ tiene la raíz asociada que hemos calculado en el ejemplo precedente, la tabla siguiente presenta un cálculo alternativo de su ciclo de formas reducidas basado en ir eliminando un coeficiente del desarrollo en fracción continua de $|\alpha|$:

α	α	$f(x,y)$
$-\overline{[2,1,1]}$	$\frac{-4-\sqrt{40}}{4}$	$2x^2+4xy-3y^2$
$\overline{[1,1,2]}$	$\frac{4+\sqrt{40}}{6}$	$-3x^2+4xy+2y^2$
$-\overline{[1,2,1]}$	$\frac{-2-\sqrt{40}}{6}$	$3x^2+2xy-3y^2$
$\overline{[2,1,1]}$	$\frac{4+\sqrt{40}}{4}$	$-2x^2+4xy+3y^2$
$-\overline{[1,1,2]}$	$\frac{-4-\sqrt{40}}{6}$	$3x^2+4xy-2y^2$
$\overline{[1,2,1]}$	$\frac{2+\sqrt{40}}{6}$	$-3x^2+2xy+3y^2$

Así ya podemos demostrar el resultado que habíamos anunciado: ∎

Teorema 11.30 *Dos formas cuadráticas reducidas de un mismo discriminante $D > 0$ son estrictamente equivalentes si y sólo si determinan el mismo ciclo de formas cuadráticas reducidas.*

Demostración: Si $f(x,y)$ y $f'(x,y)$ son dos formas reducidas de discriminante D estrictamente equivalentes, no perdemos generalidad si suponemos que sus coeficientes respectivos a, a' son negativos, pues en caso contrario podemos sustituirlas por cualquiera de sus formas adyacentes reducidas. Esto hace que sus raíces asociadas α y α' sean positivas.

El teorema 11.29 nos da que

$$\alpha = \frac{r\alpha' + s}{r'\alpha' + s'},$$

para ciertos enteros r, s, r', s' tales que $rs' - sr' = 1$. El teorema 11.15 implica entonces que α y α' (o, lo que es lo mismo, $|\alpha|$ y $|\alpha'|$) tienen desarrollos en fracción continua finalmente iguales, pero, como sus desarrollos son periódicos, esto significa que $|\alpha|$ y $|\alpha'|$ tienen el mismo periodo (tal vez desfasado, como en $\overline{[1,1,2]}$ y $\overline{[2,1,1]}$), luego ambas fracciones continuas determinan el mismo ciclo de formas cuadráticas reducidas. ∎

Ejemplo Es fácil ver que sólo hay dos formas cuadráticas reducidas de discriminante $D = 53$, que son

$$x^2 + 7xy - y^2, \qquad -x^2 + 7xy + y^2,$$

y además ambas forman un ciclo de formas estrictamente equivalentes (esto es obvio a priori, porque los ciclos tienen longitud par), así que hay una única clase de equivalencia de formas de discriminante $D = 53$. Un representante más simple es $x^2 - xy - 13y^2$.

Por lo tanto, el teorema 11.20 nos permite concluir que un primo p es de la forma $p = x^2 - xy - 13y^2$ si y sólo si $p = 53$ o $(53/p) = 1$, es decir, si y sólo si

$$p = 53 \quad \text{o} \quad \pm p \equiv 1, 4, 6, 7, 9, 10, 11, 13, 15, 16, 17, 24, 25 \text{ (mód } 53).$$

Observemos que los números de la forma $x^2 - xy - 13y^2$ son las normas de los elementos del anillo $\mathbb{Z}[\omega]$, donde $\omega = \frac{1+\sqrt{53}}{2}$, por lo que ésta es la conclusión a la que habríamos llegado en virtud del teorema 9.13 si supiéramos que este anillo es un dominio de factorización única (teniendo en cuenta que, como es fácil comprobar, la unidad fundamental es $\epsilon = 3 + \omega$ y tiene norma -1, por lo que no tenemos que preocuparnos por el signo de la norma). Sin embargo $\mathbb{Q}\big(\sqrt{53}\big)$ no está en la lista de cuerpos cuadráticos euclídeos dada por el teorema 9.11. Veremos más adelante (teorema 13.21) que si k es un cuerpo cuadrático de discriminante D y sólo hay una clase de equivalencia estricta de formas cuadráticas de dicho discriminante, entonces el anillo de enteros algebraicos de k tiene factorización única, por lo que $\mathbb{Q}\big(\sqrt{53}\big)$ es un ejemplo de cuerpo cuadrático con factorización única no euclídeo. ∎

Ejercicio: Continuando con el ejemplo anterior, probar que si $\alpha \in \mathbb{Z}[\omega]$, entonces uno de los asociados $\alpha\epsilon^i$, para cierto exponente $i = 0, 1, 2$, está en $\mathbb{Z}[\sqrt{53}]$. Caracterizar los primos de la forma $p = x^2 - 53y^2$.

Ejemplo Consideremos ahora las formas cuadráticas de discriminante $D = 56$.

$$\begin{array}{cccc} x^2 + 6xy - 5y^2 & -5x^2 + 4xy + 2y^2 & 2x^2 + 4xy - 5y^2 & -5x^2 + 6xy + y^2 \\ -x^2 + 6xy + 5y^2 & 5x^2 + 4xy - 2y^2 & -2x^2 + 4xy + 5y^2 & 5x^2 + 6xy - y^2 \end{array}$$

Es fácil ver que forman dos ciclos de longitud 4, por lo que hay dos clases de equivalencia estricta. Dos representantes más sencillos son

$$x^2 - 14y^2, \qquad 14x^2 - y^2.$$

Obviamente, ambas formas representan los mismos números salvo el signo, luego el teorema 11.20 nos asegura que un primo p es de la forma $\pm p = x^2 - 14y^2$ si y sólo si $p = 2, 7$ o $(14/p) = 1$, explícitamente, si y sólo si

$$p = 2, 7 \quad \text{o} \quad \pm p \equiv 1, 5, 9, 11, 13, 25 \text{ (mód } 56).$$

Nuevamente, éste es el mismo resultado que obtendríamos de 9.13 si supiéramos que el anillo $\mathbb{Z}\big[\sqrt{14}\,\big]$ es un dominio de factorización única. ∎

Ejercicio: Caracterizar los primos de la forma $p = x^2 - 14y^2$.

Ejemplo/ejercicio La tabla siguiente contiene todas las formas cuadráticas reducidas de discriminante $D = 328$:

$$\begin{array}{rrr}
-11x^2 + 14xy + 3y^2 & 3x^2 + 16xy + 6y^2 & -6x^2 + 8xy + 11y^2 \\
11x^2 + 14xy - 3y^2 & -3x^2 + 16xy + 6y^2 & 6x^2 + 8xy - 11y^2 \\
-11x^2 + 8xy + 6y^2 & 6x^2 + 16xy - 3y^2 & -3x^2 + 14xy + 11y^2 \\
11x^2 + 8xy - 6y^2 & -6x^2 + 16xy + 3y^2 & 3x^2 + 14xy - 11y^2 \\
-9x^2 + 16xy + 2y^2 & 2x^2 + 16xy - 9y^2 & -9x^2 + 2xy + 9y^2 \\
9x^2 + 16xy - 2y^2 & -2x^2 + 16xy + 9y^2 & 9x^2 + 2xy - 9y^2 \\
x^2 + 18xy - y^2 & -x^2 + 18xy + y^2 &
\end{array}$$

Comprobar que hay cuatro clases de equivalencia estricta, a saber,

$$\begin{array}{ll}
1 = [x^2 + 18xy - y^2], & a = [\ \ 3x^2 + 14xy - 11y^2], \\
b = [2x^2 + 16xy - 9y^2], & c = [-11x^2 + 14xy + \ 3y^2].
\end{array}$$

Es claro que a y c forman una misma clase de equivalencia no estricta, mientras que los representantes de 1 y b no son equivalentes, pues toman valores distintos.

Comprobar (empíricamente, es decir, sin tratar de justificarlo) que al multiplicar un número de una de estas cuatro formas por un número de otra, el resultado es de la forma indicada en la tabla siguiente:

	1	a	b	c
1	1	a	b	c
a	a	b	c	1
b	b	c	1	a
c	c	1	a	b

Por ejemplo, el producto de $6 = 3{\cdot}1^2 + 14{\cdot}1{\cdot}1 - 11{\cdot}1^2$ por $31 = 2{\cdot}2^2 + 16{\cdot}2{\cdot}1 - 9{\cdot}1^2$ es $186 = -11 \cdot 3^2 + 14 \cdot 3 \cdot 5 + 3 \cdot 5^2$, en correspondencia con la entrada de la tabla $a \cdot b = c$.

Notemos que, como a y c toman los mismos valores, la tabla puede resumirse así:

	1	a	b
1	1	a	b
a	a	$1/b$	a
b	b	a	1

donde $a \cdot a = a \cdot c$ es a la vez 1 y b, es decir, que el producto de un número de la forma a por otro de la forma b es a la vez de la forma 1 y b. Por ejemplo, el producto de

$$6 = 3 \cdot 1^2 + 14 \cdot 1 \cdot 1 - 11 \cdot 1^2 \quad \text{por} \quad 11 = 3 \cdot 11^2 + 14 \cdot 11 \cdot (-2) - 11 \cdot (-2)^2$$

es

$$66 = 1^2 + 18 \cdot 1 \cdot 5 - 5^2 = 2 \cdot 3^2 + 16 \cdot 3 \cdot 4 - 9 \cdot 4^2.$$

Sin embargo, expresando así la tabla perdemos la estructura de grupo cíclico que muestra la primera tabla, en la que se distinguen las clases a y c, a pesar de que corresponden a una misma clase de equivalencia y representan, por tanto, los mismos números. En el capítulo siguiente podremos justificar estos hechos.

∎

Ejercicio: Repetir el ejercicio anterior con las formas reducidas de discriminante $D = 316$:

$x^2 + 16xy - 15y^2$	$-15x^2 + 14xy + 2y^2$	$2x^2 + 14xy - 15y^2$	$-15x^2 + 16xy + y^2$
$15x^2 + 16xy - y^2$	$-x^2 + 16xy + 15y^2$	$15x^2 + 14xy - 2y^2$	$-2x^2 + 14xy + 15y^2$
$10x^2 + 14xy - 3y^2$	$-3x^2 + 16xy + 5y^2$	$5x^2 + 14xy - 6y^2$	$-6x^2 + 10xy + 9y^2$
$9x^2 + 8xy - 7y^2$	$-7x^2 + 6xy + 10y^2$	$10x^2 + 6xy - 7y^2$	$-7x^2 + 8xy + 9y^2$
$9x^2 + 10xy - 6y^2$	$-6x^2 + 14xy + 5y^2$	$5x^2 + 16xy - 3y^2$	$-3x^2 + 14xy + 10y^2$
$-10x^2 + 6xy + 7y^2$	$7x^2 + 8xy - 9y^2$	$-9x^2 + 10xy + 6y^2$	$6x^2 + 14xy - 5y^2$
$-5x^2 + 16xy + 3y^2$	$3x^2 + 14xy - 10y^2$	$-9x^2 + 8xy + 7y^2$	$7x^2 + 6xy - 10y^2$
$-10x^2 + 14xy + 3y^2$	$3x^2 + 16xy - 5y^2$	$-5x^2 + 14xy + 6y^2$	$6x^2 + 10xy - 9y^2$

Ejercicio: Probar que toda forma cuadrática de discriminante $D = 60$ es estrictamente equivalente a una de las cuatro formas siguientes:

$$x^2 - 15y^2, \qquad 15x^2 - y^2, \qquad 3x^2 - 5y^2, \qquad 5x^2 - 3y^2.$$

Capítulo XII

Módulos

En el capítulo anterior hemos planteado el problema de encontrar un método para resolver las ecuaciones diofánticas determinadas por formas cuadráticas de dos variables. Ya sabemos cómo resolver las asociadas a formas reducibles, y aquí vamos a abordar el caso en que la forma es irreducible. Sabemos que encontrar las soluciones enteras de la ecuación $x^2+y^2 = 65$ equivale a encontrar los enteros de Gauss de norma 65, y que este enfoque, que en principio podría parecer rebuscado, resulta ser el más conveniente para entender y resolver el problema (pues nos permite aprovechar la factorización única de los enteros de Gauss). Ahora generalizaremos la idea al caso de formas cuadráticas irreducibles (primitivas, etc.) arbitrarias.

12.1 Módulos en cuerpos cuadráticos

Consideremos por ejemplo una ecuación como

$$2x^2 + 22xy - 7y^2 = 77.$$

El miembro izquierdo es una forma cuadrática de discriminante 540. Las raíces del polinomio $2x^2 + 22x - 7$ son

$$\alpha = \frac{-11 - 3\sqrt{15}}{2}, \qquad \bar{\alpha} = \frac{-11 + 3\sqrt{15}}{2},$$

donde la barra indica la conjugación en el cuerpo $\mathbb{Q}\big(\sqrt{15}\,\big)$. Por lo tanto,

$$2x^2 + 22x - 7 = 2(x - \alpha)(x - \bar{\alpha})$$

y cambiando x por x/y llegamos a que

$$2x^2 + 22xy - 7y^2 = 2(x - \alpha y)(x - \bar{\alpha} y) = 2\,\mathrm{N}(x - \alpha y).$$

Por lo tanto, si llamamos M al conjunto de todos los elementos de $\mathbb{Q}\big(\sqrt{15}\,\big)$ de la forma $x-\alpha y$, donde x, y son enteros, tenemos que las soluciones de la ecuación dada se corresponden con los elementos de M de norma $77/2$. Nuevamente

este enfoque puede parecer rebuscado, pero veremos que, si lo desarrollamos convenientemente, nos proporcionará un método para resolver cualquier ecuación diofántica definida por una forma cuadrática irreducible.

Definición 12.1 Sea $k = \mathbb{Q}(\sqrt{d}\,)$ un cuerpo cuadrático. El *módulo* de k generado por los elementos $\alpha_1, \ldots, \alpha_r$ de k es el conjunto $M = \langle \alpha_1, \ldots, \alpha_r \rangle$ formado por los elementos de k de la forma

$$x_1\alpha_1 + \cdots + x_r\alpha_r,$$

donde $x_1, \ldots, x_r$ son enteros racionales. Se dice que $\alpha_1, \ldots, \alpha_r$ son un *sistema generador* de M.

Acabamos de ver que resolver la ecuación diofántica $2x^2 + 22xy - 7y^2 = 77$ equivale a encontrar los elementos del módulo $M = \langle 1, -\alpha \rangle$ de norma $77/2$.

En general, si tenemos cualquier ecuación diofántica $ax^2 + bxy + cy^2 = e$, donde el miembro izquierdo es una forma cuadrática (primitiva) de discriminante $D = m^2 d$, con d libre de cuadrados, podemos factorizar

$$ax^2 + bx + c = a(x - \alpha)(x - \bar{\alpha}),$$

donde

$$\alpha = \frac{-b + m\sqrt{d}}{2a}.$$

De aquí podemos pasar a que

$$ax^2 + bxy + cy^2 = a\,\mathrm{N}(x - \alpha y),$$

luego las soluciones de la ecuación se corresponden con los elementos del módulo $M = \langle 1, -\alpha \rangle$ de norma e/a.

Generadores y bases Vamos a estudiar los módulos con más detenimiento. En primer lugar vamos a probar que todo módulo admite un sistema generador con a lo sumo dos elementos. Para ello conviene aislar un hecho elemental que usaremos en diversas ocasiones:

> *Si en un sistema generador de un módulo sumamos a uno de sus elementos un múltiplo de otro, obtenemos otro sistema generador del mismo módulo.*

Equivalentemente:

$$\langle \alpha_1, \ldots, \alpha_r \rangle = \langle \alpha_1 + k\alpha_2, \alpha_2, \ldots, \alpha_r \rangle,$$

para todo entero k.

En efecto, todo elemento del módulo de la izquierda es de la forma

$$x_1\alpha_1 + x_2\alpha_2 + \cdots + x_r\alpha_r = x_1(\alpha_1 + k\alpha_2) + (x_2 - x_1k)\alpha_2 + \cdots + x_r\alpha_r,$$

luego todo elemento del módulo de la izquierda está en el de la derecha y, recíprocamente:

$$x_1(\alpha_1 + k\alpha_2) + x_2\alpha_2 + \cdots + x_r\alpha_r = x_1\alpha_1 + (x_2 + x_1k)\alpha_2 + \cdots + x_r\alpha_r,$$

luego todo elemento del módulo de la derecha está en el de la izquieda. ∎

También es útil tener en cuenta que si en un sistema generador de un módulo cambiamos un elemento α por $-\alpha$, obtenemos claramente otro generador del mismo módulo.

Ahora ya podemos probar:

Teorema 12.2 *Todo módulo M de un cuerpo cuadrático $\mathbb{Q}(\sqrt{d})$ admite un generador con a lo sumo dos elementos.*

Demostración: Vamos a dar un algoritmo explícito para reducir los generadores de un módulo dado. Lo explicamos con un ejemplo, pero será obvio que el procedimiento vale en general. Consideremos por ejemplo el módulo $M = \langle \alpha_1, \alpha_2, \alpha_3, \alpha_4 \rangle$, donde

$$\alpha_1 = \frac{6}{5} + \frac{2}{15}\sqrt{d}, \quad \alpha_2 = -\frac{3}{10} + \frac{29}{30}\sqrt{d}, \quad \alpha_3 = \frac{3}{5} + \frac{9}{10}\sqrt{d}, \quad \alpha_4 = \frac{1}{2} - \frac{1}{6}\sqrt{d}.$$

En primer lugar expresamos todas las fracciones de cada coordenada con el mismo denominador:

$$\alpha_1 = \frac{12}{10} + \frac{4}{30}\sqrt{d},\ \alpha_2 = -\frac{3}{10} + \frac{29}{30}\sqrt{d},\ \alpha_3 = \frac{6}{10} + \frac{27}{30}\sqrt{d},\ \alpha_4 = \frac{5}{10} - \frac{5}{30}\sqrt{d}.$$

Equivalentemente:

$$\alpha_1 = 12\gamma + 4\delta, \quad \alpha_2 = -3\gamma + 29\delta, \quad \alpha_3 = 6\gamma + 27\delta, \quad \alpha_4 = 5\gamma + 5\delta,$$

donde $\gamma = 1/10$, $\delta = \sqrt{d}/30$.

Ahora sólo tenemos que usar sistemáticamente la observación precedente al enunciado. Para agilizar los cálculos podemos representar abreviadamente los cuatro generadores que tenemos como

$$\begin{pmatrix} 12 & 4 \\ -3 & 29 \\ 5 & 5 \\ 6 & 27 \end{pmatrix}$$

Hemos puesto en primer lugar el generador cuya segunda coordenada es menor en valor absoluto. Dividimos $29 = 4 \cdot 7 + 1$ y sumamos al segundo generador el primero multiplicado por -7. Igualmente, al tercero le sumamos el primero multiplicado por -1 y al cuarto el primero multiplicado por -6. El resultado es la primera matriz de la sucesión siguiente:

$$\begin{pmatrix} 12 & 4 \\ -87 & 1 \\ -66 & 3 \\ -7 & 1 \end{pmatrix} \sim \begin{pmatrix} -7 & 1 \\ 40 & 0 \\ -80 & 0 \\ -45 & 0 \end{pmatrix} \sim \begin{pmatrix} -7 & 1 \\ 40 & 0 \\ 0 & 0 \\ -5 & 0 \end{pmatrix} \sim \begin{pmatrix} 3 & 1 \\ 5 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Hemos puesto $(-7,1)$ en primer lugar porque tiene su segunda componente con menor valor absoluto, y a las otras tres filas les hemos sumado el múltiplo adecuado de la primera para reducir la segunda componente. En este caso todas se han hecho nulas. En general, como en cada paso reducimos las segundas componentes, tras un número finito de pasos se tiene que llegar a lo que hemos obtenido: que todas las segundas componentes son nulas salvo a lo sumo la de la primera fila. Ahora usamos la primera componente de $(40,0)$ para reducir las primeras componentes de las demás filas. Ponemos en primer lugar la de menor valor absoluto, que es $(-5,0)$ (pero le cambiamos el signo por simplicidad) y con ese 5 reducimos las primeras componentes de los restantes. Tras un número finito de pasos, todas se tienen que hacer todas nulas salvo a lo sumo las dos primeras (la primera no podemos hacerla nula porque no podemos usar la primera fila para reducir las demás, ya que entonces desharíamos los ceros que hemos hecho en las segundas componentes). Finalmente, es claro que podemos eliminar los generadores nulos, y en nuestro caso concluimos que

$$M = \langle 5\gamma, 3\gamma + \delta\rangle = \left\langle \frac{1}{2}, \frac{3}{10} + \frac{\sqrt{d}}{30} \right\rangle.$$

Notemos que si todos los generadores tuvieran nula su segunda coordenada, llegaríamos a un único generador final. ∎

Ahora es natural preguntarse si un módulo con dos generadores puede expresarse también en términos de un único generador:

Teorema 12.3 *Sea $M = \langle \alpha, \beta\rangle$ un módulo en un cuerpo cuadrático $\mathbb{Q}(\sqrt{d}\,)$, donde*

$$\alpha = r + s\sqrt{d}, \qquad \beta = r' + s'\sqrt{d}.$$

Entonces M puede generarse mediante un único elemento si y sólo si

$$\begin{vmatrix} r & r' \\ s & s' \end{vmatrix} = 0.$$

Demostración: Sea m el mínimo común múltiplo de los denominadores de r, s, r', s', de modo que

$$r = \frac{r_0}{m}, \quad s = \frac{s_0}{m}, \quad r' = \frac{r'_0}{m}, \quad s' = \frac{s'_0}{m},$$

donde r_0, s_0, r'_0, s'_0 son enteros.

Es inmediato comprobar que el determinante del enunciado vale 0 si y sólo si

$$\begin{vmatrix} r_0 & r'_0 \\ s_0 & s'_0 \end{vmatrix} = 0.$$

por el teorema 11.9 esto equivale a que $(r_0, s_0) = k(u,v)$, $(r'_0, s'_0) = l(u,v)$, donde $(k,l) = 1$.

Por el teorema de Bezout existen enteros tales que $ak + bl = 1$, luego

$$a\alpha + b\beta = a\left(\frac{r_0}{m} + \frac{s_0}{m}\sqrt{d}\right) + b\left(\frac{r'_0}{m} + \frac{s'_0}{m}\sqrt{d}\right) =$$

$$(ak + bl)\left(\frac{u}{m} + \frac{v}{m}\sqrt{d}\right) = \frac{u}{m} + \frac{v}{m}\sqrt{d},$$

luego $\frac{u}{m} + \frac{v}{m}\sqrt{d}$ es un elemento de M, y es un generador, pues todo elemento de M es de la forma

$$x\alpha + y\beta = x\left(\frac{r_0}{m} + \frac{s_0}{m}\sqrt{d}\right) + y\left(\frac{r'_0}{m} + \frac{s'_0}{m}\sqrt{d}\right) =$$

$$(xk + vl)\left(\frac{u}{m} + \frac{v}{m}\sqrt{d}\right).$$

Recíprocamente, si $M = \left\langle u + v\sqrt{d}\right\rangle$, entonces existen enteros a y b tales que

$$\alpha = a(u + v\sqrt{d}), \quad \beta = b(u + v\sqrt{d}),$$

luego

$$\begin{vmatrix} r & r' \\ s & s' \end{vmatrix} = \begin{vmatrix} au & bu \\ av & bv \end{vmatrix} = 0.$$

∎

Definición 12.4 Un módulo M en un cuerpo cuadrático $\mathbb{Q}(\sqrt{d}\,)$ es *completo* si no admite sistemas generadores con un único elemento.

El teorema anterior nos da un criterio sencillo para determinar si un módulo es completo, pero podemos expresarlo de otra forma aún más conveniente. Para ello, si α, β son dos elementos de $\mathbb{Q}(\sqrt{d}\,)$, definimos el *discriminante orientado* de α, β como

$$\Delta_0[\alpha, \beta] = \begin{vmatrix} \bar{\alpha} & \bar{\beta} \\ \alpha & \beta \end{vmatrix}.$$

Este discriminante tiene una propiedad fundamental:

Si α, β son dos elementos de un cuerpo cuadrático $\mathbb{Q}(\sqrt{d}\,)$ y

$$\gamma = r\alpha + s\beta, \qquad \delta = r'\alpha + s'\beta,$$

para ciertos números racionales r, s, r', s', entonces

$$\Delta_0[\gamma, \delta] = \begin{vmatrix} r & r' \\ s & s' \end{vmatrix} \Delta_0[\alpha, \beta].$$

En efecto, basta observar que

$$\begin{pmatrix} \bar{\gamma} & \bar{\delta} \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} r\bar{\alpha} + s\bar{\beta} & r'\bar{\alpha} + s'\bar{\beta} \\ r\alpha + s\beta & r'\alpha + s'\beta \end{pmatrix} = \begin{pmatrix} \bar{\alpha} & \bar{\beta} \\ \alpha & \beta \end{pmatrix} \begin{pmatrix} r & r' \\ s & s' \end{pmatrix},$$

y tomar determinantes.

En particular, si $\alpha = r + s\sqrt{d}$ y $\beta = r' + s'\sqrt{d}$, tenemos que

$$\Delta_0[\alpha,\beta] = \begin{vmatrix} r & r' \\ s & s' \end{vmatrix} \Delta_0[1,\sqrt{d}],$$

pero

$$\Delta_0[1,\sqrt{d}] = \begin{vmatrix} 1 & -\sqrt{d} \\ 1 & \sqrt{d} \end{vmatrix} = 2\sqrt{d} \neq 0,$$

luego el teorema anterior puede reformularse en estos términos:

Un módulo $M = \langle\alpha,\beta\rangle$ es completo si y sólo si $\Delta_0[\alpha,\beta] \neq 0$.

Más aún, si γ,δ son dos elementos cualesquiera de M, se cumple que

$$\gamma = r\alpha + s\beta, \qquad \delta = r'\alpha + s'\beta$$

con r, s, r', s' enteros, por lo que $\Delta_0[\gamma,\delta] = k\Delta_0[\alpha,\beta]$, donde k es entero (es el determinante de una matriz con coeficientes enteros). Si γ,δ es otro sistema generador de M, entonces se cumple también que $\Delta_0[\alpha,\beta] = l\Delta_0[\gamma,\delta]$, para cierto entero k, de modo que $kl = 1$, luego $k = l = \pm 1$. Equivalentemente:

Si α,β y γ,δ son dos sistemas generadores de un mismo módulo M, entonces $\Delta_0[\alpha,\beta] = \pm\Delta_0[\gamma,\delta]$.

Para evitar esta variación de signo, así como el hecho obvio de que

$$\Delta_0[\alpha,\beta] = -\Delta_0[\beta,\alpha],$$

definimos el *discriminante* (no orientado) de α,β como

$$\Delta[\alpha,\beta] = \Delta_0[\alpha,\beta]^2.$$

Este discriminante tiene una propiedad adicional, y es que es siempre un número racional.

En efecto, de la definición de $\Delta_0[\alpha,\beta]$ se sigue inmediatamente que

$$\overline{\Delta_0[\alpha,\beta]} = \Delta_0[\bar\alpha,\bar\beta] = -\Delta_0[\alpha,\beta],$$

luego

$$\overline{\Delta[\alpha,\beta]} = (\overline{\Delta_0[\alpha,\beta]})^2 = (-\Delta_0[\alpha,\beta])^2 = \Delta[\alpha,\beta]$$

y los únicos elementos de k que son iguales a sus conjugados son los de $\mathbb{Q}$.

El teorema siguiente recoge los hechos que hemos probado en términos del discriminante:

Teorema 12.5 *Sea $M = \langle\alpha,\beta\rangle$ un módulo en un cuerpo cuadrático $\mathbb{Q}(\sqrt{d})$. Entonces:*

1. *M es completo si y sólo si $\Delta[\alpha,\beta] \neq 0$.*
2. *Si $M = \langle\gamma,\delta\rangle$, entonces $\Delta[\gamma,\delta] = \Delta[\alpha,\beta]$.*

Definición 12.6 Una *base* de un módulo completo M es un generador de M formado por dos elementos. Definimos el *discriminante* de un módulo completo M como el discriminante $\Delta[M]$ de cualquiera de sus bases.

Ejemplo Si $k = \mathbb{Q}(\sqrt{d}\,)$ es un cuerpo cuadrático, su anillo de enteros $\mathcal{O} = \mathbb{Z}[\omega]$, donde $\omega = \sqrt{d}$ o bien $\omega = (1+\sqrt{d})/2$ es un módulo completo de k con base $1, \omega$ y su discriminante es

$$\Delta[1,\omega] = \begin{vmatrix} 1 & \bar{\omega} \\ 1 & \omega \end{vmatrix}^2 = (\omega - \bar{\omega})^2 = \begin{cases} 4d & \text{si } d \not\equiv 1 \pmod 4, \\ d & \text{si } d \equiv 1 \pmod 4. \end{cases}$$

Así pues:

> *El discriminante Δ_k de un cuerpo cuadrático es el discriminante de su anillo de enteros algebraicos, en el sentido de la definición precedente.*[1]

Más aún, en la sección 10.4 definimos los órdenes cuadráticos $\mathcal{O}_m$ de $\mathbb{Q}(\sqrt{d}\,)$ formados por los elementos de la forma $a + bm\omega$, donde a y b son enteros racionales. También es inmediato que $\mathcal{O}_m = \langle 1, m\omega\rangle$, así como que se trata de un módulo completo de discriminante $\Delta[1, m\omega] = m^2\Delta_k$. ∎

Discriminantes de órdenes cuadráticos Conviene observar que un orden cuadrático está completamente determinado por su discriminante, es decir, que si dos órdenes cuadráticos tienen el mismo discriminante es que se trata exactamente del mismo orden del mismo cuerpo cuadrático.

En efecto, consideremos el orden $\mathcal{O}_m$ de un cuerpo $k = \mathbb{Q}(\sqrt{d}\,)$ y el orden $\mathcal{O}_{m'}$ de un cuerpo $k' = \mathbb{Q}(\sqrt{d'}\,)$ y supongamos que $m^2\Delta_k = m'^2\Delta_{k'}$. Basta observar que tanto si $\Delta_k = d$ como si $\Delta_k = 4d$, se cumple que d es la parte libre de cuadrados de Δ_k, luego también la de $m^2\Delta_k$, luego la igualdad de los discriminantes implica que $d = d'$ (es decir, $k = k'$), luego $\Delta_k = \Delta_{k'}$, luego $m = m'$. ∎

Vamos a ver que los módulos arbitrarios se parecen —hasta cierto punto— a los módulos particulares que habíamos estudiado hasta ahora (los órdenes cuadráticos). Por ejemplo, sabemos que cada elemento de $\mathbb{Q}(i)$ se expresa de forma única como $r+si$, donde r, s son números racionales, y que cada elemento de $\mathbb{Z}[i]$ se expresa de forma única como $a+bi$, donde a, b son enteros racionales. Esto es un caso particular del teorema siguiente:

Teorema 12.7 *Sea $k = \mathbb{Q}(\sqrt{d}\,)$ un cuerpo cuadrático y $M = \langle\alpha, \beta\rangle$ un módulo en k. Entonces cada elemento de k se expresa de forma única como $\gamma = r\alpha + s\beta$, donde r, s son números racionales llamados* coordenadas *de γ en la base α, β. Así, los elementos de M son los elementos de k con coordenadas enteras.*

[1] Esto prueba lo que habíamos afirmado en la primera nota al pie de la página 301, es decir, que si dos cuerpos cuadráticos $\mathbb{Q}(\sqrt{d})$ y $\mathbb{Q}(\sqrt{d'})$ fueran iguales, entonces tendrían el mismo discriminante.

DEMOSTRACIÓN: Pongamos que $\alpha = r+s\sqrt{d}$, $\beta = r'+s'\sqrt{d}$, donde r, s, r', s' son números racionales. Esto puede expresarse matricialmente:

$$(\alpha, \beta) = (1, \sqrt{d}) \begin{pmatrix} r & r' \\ s & s' \end{pmatrix}.$$

Pero hemos visto que el hecho de que M sea un módulo completo equivale a que la matriz tenga determinante no nulo, luego tiene inversa en $\mathrm{Mat}_2(\mathbb{Q})$, es decir, existe otra matriz tal que

$$\begin{pmatrix} r & r' \\ s & s' \end{pmatrix} \begin{pmatrix} u & u' \\ v & v' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Multiplicando por la matriz inversa la primera ecuación matricial resulta que

$$(\alpha, \beta) \begin{pmatrix} u & u' \\ v & v' \end{pmatrix} = (1, \sqrt{d}).$$

Explícitamente, $1 = u\alpha + v\beta$, $\sqrt{d} = u'\alpha + v'\beta$, y un elemento arbitrario de k es de la forma

$$\gamma = a + b\sqrt{d} = a(u\alpha + v\beta) + b(u'\alpha + v'\beta) = (au + bu')\alpha + (av + bv')\beta.$$

Veamos ahora que la expresión es única. En caso contrario, tendríamos que

$$\gamma = r\alpha + s\beta = r'\alpha + s'\beta$$

con $(r, s) \neq (r', s')$. Supongamos, por ejemplo, que $r \neq r'$, Entonces tenemos que $(r - r')\alpha = (s' - s)\beta$ y $\alpha = t\beta$, para cierto número racional t, pero esto hace que

$$\Delta[\alpha, \beta] = \begin{vmatrix} \bar{\alpha} & t\bar{\alpha} \\ \alpha & t\alpha \end{vmatrix}^2 = (t\alpha\bar{\alpha} - t\alpha\bar{\alpha})^2 = 0,$$

contradicción. ∎

A la vista del teorema anterior, es razonable definir una *base* de un cuerpo cuadrático k como cualquier par de elementos α, β de k tales que $\Delta[\alpha, \beta] \neq 0$, pues cuando se cumple esta condición sabemos que todo elemento de k se expresa de forma única como $r\alpha + s\beta$, para ciertos números racionales r y s.

En estos términos, cada base de k es también una base del módulo completo $M = \langle \alpha, \beta \rangle$ que genera, formado por los elementos de k con coordenadas enteras.

Coeficientes Volvamos a la relación de todo esto con las ecuaciones diofánticas definidas por formas cuadráticas. Hemos visto que toda forma cuadrática irreducible de discriminante $D = m^2 d$ factoriza como

$$ax^2 + bxy + cy^2 = a\,\mathrm{N}(x - \alpha y),$$

donde

$$\alpha = \frac{-b + m\sqrt{d}}{2a},$$

por lo que el discriminante del módulo $M = \langle 1, -\alpha \rangle$ es

$$\Delta[M] = \begin{vmatrix} 1 & \frac{b+m\sqrt{d}}{2a} \\ 1 & \frac{b-m\sqrt{d}}{2a} \end{vmatrix}^2 = \left(\frac{-m\sqrt{d}}{a}\right)^2 = \frac{D}{a^2}.$$

En particular vemos que M es siempre un módulo completo.

Ejemplo Consideremos de nuevo la ecuación diofántica

$$2x^2 + 22xy - 7y^2 = 77$$

Una solución es $(x, y) = (3, 1)$. Más adelante veremos cómo podemos encontrarla sin recurrir a una mera exploración, pero de momento vamos a ver que a partir de esta solución podemos construir muchas más. Habíamos visto que la ecuación puede escribirse como

$$2\,\mathrm{N}(x - \alpha y) = 77,$$

donde

$$\alpha = \frac{-11 - 3\sqrt{15}}{2}.$$

Equivalentemente, las soluciones enteras de la ecuación se corresponden con los elementos del módulo

$$M = \left\langle 1, \frac{11 + 3\sqrt{15}}{2} \right\rangle$$

de norma $77/2$. La solución que hemos dado corresponde a

$$\xi = 3 - \alpha = \frac{17 + 3\sqrt{15}}{2}.$$

Consideremos ahora $\epsilon = 244 + 63\sqrt{15}$, que tiene norma 1. Si calculamos

$$\epsilon\xi = \frac{6\,983 + 1\,803\sqrt{15}}{2} = 186 + 601\,\frac{11 + 3\sqrt{15}}{2}$$

vemos que $\epsilon\xi$ es también un elemento de M, que necesariamente tendrá norma $77/2$, luego determina otra solución de la ecuación, a saber $(x, y) = (186, 601)$.

Ejercicio: Comprobar que $\epsilon^2\xi$ proporciona otra solución de la ecuación.

Aquí no sólo es esencial que $\mathrm{N}(\epsilon) = 1$, sino también que $\epsilon\xi$ está en M. Por ejemplo, si tratamos de usar $\epsilon = 4 + \sqrt{15}$, que también tiene norma 1 (es la unidad fundamental de $\mathbb{Q}(\sqrt{15})$), obtenemos

$$\epsilon\xi = \frac{113 + 29\sqrt{15}}{2} = \frac{20}{6} + \frac{29}{3}\,\frac{11 + 3\sqrt{15}}{2},$$

que no está en M, pues tiene coordenadas fraccionarias respecto de la base que estamos considerando. Así sólo obtenemos la solución racional, pero no entera, $(x, y) = (20/6, 29/3)$.

De este modo, el proceso seguido para obtener la solución $(186, 601)$ a partir de $(3, 1)$ plantea un interrogante mayor que lo que aporta a la hora de resolver la ecuación: ¿De dónde hemos sacado un valor tan oportuno para ϵ? Vamos a introducir algunos conceptos más que nos ayudarán a responder esta pregunta:

Definición 12.8 Si M es un módulo completo en un cuerpo cuadrático k y $\gamma \neq 0$ es un elemento de k, definimos γM como el conjunto de los elementos $\gamma\delta$, tales que δ varía en M.

Se cumple que γM es un módulo completo. En efecto, si $M = \langle \alpha, \beta \rangle$, es claro que $\gamma M = \langle \gamma\alpha, \gamma\beta \rangle$, y

$$\Delta_0[\gamma\alpha, \gamma\beta] = \begin{vmatrix} \bar{\gamma}\bar{\alpha} & \bar{\gamma}\bar{\beta} \\ \gamma\alpha & \gamma\beta \end{vmatrix} = \mathrm{N}(\gamma)\bar{\alpha}\beta - \mathrm{N}(\gamma)\alpha\bar{\beta} = \mathrm{N}(\gamma)\Delta_0[\alpha, \beta].$$

En particular,

$$\Delta[\gamma M] = \mathrm{N}(\gamma)^2\Delta[M].$$

Un *coeficiente* de M es un elemento γ de k tal que $\gamma M \subset M$. El *anillo de coeficientes* de M es el conjunto $\mathcal{O}_M$ de todos los coeficientes de M.

Es fácil comprobar que, ciertamente, se trata de un subanillo de k:

1. *Si α y β son coeficientes de M, también lo son $\alpha \pm \beta$.*

   En efecto, si γ es un elemento de M, entonces $\alpha\gamma$ y $\beta\gamma$ están en $\alpha M \subset M$ y $\beta M \subset M$, respectivamente, luego están en M y lo mismo sucede con $\alpha\gamma \pm \beta\gamma = (\alpha \pm \beta)\gamma$, luego $(\alpha \pm \beta)M \subset M$.

2. *Si α y β son coeficientes de M, también lo es $\alpha\beta$.*

   En efecto, si γ es un elemento de M, también lo es $\beta\gamma$ (porque β es un coeficiente) y, a su vez, también lo es $\alpha\beta\gamma$ (porque α es un coeficiente), luego $\alpha\beta M \subset M$.

   También es obvio que 0 y 1 son coeficientes de M, luego $\mathcal{O}_M$ es un anillo unitario.

En estos términos, podemos decir que los valores de ϵ que nos permiten encontrar más soluciones de una ecuación diofántica asociada a un módulo M a partir de una dada son los coeficientes de M de norma 1. Vamos a ver que es fácil determinar el anillo de coeficientes de un módulo dado y que, a su vez, encontrar sus elementos de norma 1, no es que sea algo obvio, pero es un problema que ya tenemos resuelto.

Para ello necesitamos un par de observaciones elementales. Una es que dos módulos completos M y γM tienen los mismos coeficientes.

En efecto, si α es un coeficiente de γM y β es un elemento de M, entonces $\gamma\beta$ está en γM, luego $\alpha\gamma\beta$ está en $\alpha\gamma M \subset \gamma M$, luego existe un β' en M tal que $\alpha\gamma\beta = \gamma\beta'$, luego $\alpha\beta = \beta'$, luego $\alpha\beta$ está en M, y esto prueba que $\alpha M \subset M$, es decir, que α es un coeficiente de M. El recíproco se prueba más fácilmente.

La segunda observación es que $\langle\alpha,\beta\rangle = \alpha\,\langle 1,\beta/\alpha\rangle$, por lo que sólo necesitamos encontrar los anillos de coeficientes de los módulos de la forma $\langle 1,\gamma\rangle$. Notemos que γ tiene que ser irracional para que el módulo sea completo.

Teorema 12.9 *Sea γ una raíz irracional de un polinomio ax^2+bx+c con coeficientes enteros primos entre sí y con $a>0$. Entonces el anillo de coeficientes del módulo $M=\langle 1,\gamma\rangle$ es $\mathcal{O}=\langle 1,a\gamma\rangle$ y $\Delta[\mathcal{O}]=b^2-4ac$.*

Demostración: Si llamamos $D=b^2-4ac=m^2d$, entonces

$$\gamma = \frac{-b\pm m\sqrt{d}}{2a}$$

está en el cuerpo cuadrático $k=\mathbb{Q}\big(\sqrt{d}\,\big)$. Todo elemento de k es de la forma $\delta=x+y\gamma$, donde x,y son números racionales. Es claro que δ es un coeficiente de M si y sólo si δ y $\delta\gamma$ están en M. Que δ esté en M equivale a que x, y sean números enteros. Por otra parte,

$$\delta\gamma = x\gamma+y\gamma^2 = x\gamma+y\frac{-b\gamma-c}{a} = -\frac{cy}{a}+\left(x-\frac{by}{a}\right)\gamma,$$

luego $\delta\gamma$ está en M si y sólo si cy/a y by/a son enteros. Como a,b,c son primos entre sí, esto equivale a que $a\mid y$ (si $p^e\mid a$, entonces p^e no divide a b o a c, luego divide a y). En definitiva, los coeficientes de M son los elementos de $\langle 1,a\gamma\rangle$. Además:

$$\Delta[\mathcal{O}] = \begin{vmatrix} 1 & a\bar\gamma \\ 1 & a\gamma \end{vmatrix}^2 = a^2(\gamma-\bar\gamma)^2 = a^2\left(\frac{\pm 2m\sqrt{d}}{2a}\right)^2 = m^2d = b^2-4ac.$$

■

En la situación del teorema anterior podemos precisar algo más: puesto que $a\gamma^2+b\gamma+c=0$, también $(a\gamma)^2+b(a\gamma)+ac=0$, luego $a\gamma$ es raíz del polinomio $x^2+bx+ac$, cuyo discriminante es también $D=b^2-4ac=m^2d$. Esto significa que $a\gamma$ es un entero algebraico del cuerpo $k=\mathbb{Q}\big(\sqrt{d}\,\big)$, luego es de la forma $a\gamma=l+k\omega$, donde l y k son enteros racionales.

Por lo tanto, $\mathcal{O}_M=\langle 1,l+k\omega\rangle=\langle 1,k\omega\rangle=\mathcal{O}_k$, donde hemos usado que —según hemos probado antes del teorema 12.2— a un generador de un módulo podemos sumarle un múltiplo de otro. Teniendo en cuenta que $\Delta[\mathcal{O}_M]=m^2d$, concluimos que $k=m$ si $d\equiv 1\pmod 4$ y $k=m/2$ en caso contrario. En particular:

Teorema 12.10 *Los anillos de coeficientes de los módulos de un cuerpo cuadrático k son los órdenes cuadráticos $\mathcal{O}_m$ de k.*

Observemos en particular que si $\mathcal{O}_m$ es un orden de un cuerpo cuadrático k, entonces es su propio anillo de coeficientes, ya que es de la forma $\mathcal{O}_m=\langle 1,m\omega\rangle$, donde $m\omega$ es un entero algebraico, luego su polinomio mínimo cumple $a=1$, luego el anillo de coeficientes de $\mathcal{O}_m$ es $\mathcal{O}_m$ por el teorema anterior.

Por otra parte, los elementos de norma 1 de un anillo $\mathcal{O}_m$ son sus unidades de norma 1, y en el capítulo anterior hemos visto cómo encontrarlas todas.

Ejemplo Continuamos con el ejemplo de la ecuación

$$2x^2 + 22xy - 7y^2 = 77,$$

cuyas soluciones se corresponden con los elementos del módulo

$$M = \left\langle 1, \frac{11 + 3\sqrt{15}}{2} \right\rangle$$

de norma 77/2. El teorema 12.9 nos dice que

$$\mathcal{O}_M = \left\langle 1, 11 + 3\sqrt{15} \right\rangle = \left\langle 1, 3\sqrt{15} \right\rangle = \mathcal{O}_3 \subset \mathbb{Q}\big(\sqrt{15}\,\big),$$

luego los coeficientes de M de norma 1 son las potencias de la unidad fundamental $\epsilon = 244 + 63\sqrt{15}$ de $\mathcal{O}_3$ (porque tiene norma 1, pues en caso contrario tendríamos que considerar ϵ^2), que sabemos calcular como hemos visto en la sección 10.4.

Por otro lado, observemos que podemos simplificar la situación teniendo en cuenta que los elementos de M de norma 77/2 se corresponden con los elementos de $2M$ de norma $77 \cdot 2$ (pues $\mathrm{N}(2) = 4$). Tenemos que

$$2M = \left\langle 2, 11 + 3\sqrt{15} \right\rangle \subset \mathcal{O}_3.$$

Además,

$$\mathrm{N}(2x + (11 + 3\sqrt{15})y) = \mathrm{N}(2)\,\mathrm{N}\left(x + \frac{11 + 3\sqrt{15}}{2}y\right) = 4x^2 + 44xy - 14y^2,$$

luego

$$2x^2 + 22xy - 7y^2 = \frac{1}{2}\,\mathrm{N}(2x + (11 + 3\sqrt{15})y)$$

y las soluciones de la ecuación se corresponden con los elementos de $2M$ de norma 154. La ventaja de trabajar con $2M$ es que sus elementos forman parte del anillo de coeficientes. Por ejemplo, demostraremos que en un módulo contenido en su anillo de coeficientes sólo hay un número finito de elementos no asociados dos a dos de una misma norma, lo cual significa que todas las soluciones de una ecuación determinada por una forma cuadrática pueden obtenerse a partir de un número finito de elementos de M multiplicándolos por unidades de su anillo de coeficientes. ∎

Ejemplo Vamos a encontrar las soluciones enteras de la ecuación

$$x^2 - 9xy - y^2 = \pm 1.$$

Las raíces del polinomio $x^2 - 9x - 1$ son

$$\epsilon = \frac{9+\sqrt{85}}{2} = 4 + \omega, \qquad \bar{\epsilon} = \frac{9-\sqrt{85}}{2} = 5 - \omega.$$

Por lo tanto,

$$x^2 - 9xy - y^2 = (x + y(-5+\omega))(x + y(-4-\omega)) = \mathrm{N}(x + y(-5+\omega)).$$

Por lo tanto, las soluciones enteras de la ecuación se corresponden con los elementos del módulo

$$M = \langle 1, -\bar{\epsilon}\rangle = \langle 1, -5+\omega\rangle = \langle 1, \omega\rangle = \mathbb{Z}[\omega]$$

de norma ± 1, es decir, con las unidades del anillo de enteros algebraicos de $\mathbb{Q}\big(\sqrt{85}\,\big)$. Notemos que $\bar{\epsilon}$ se corresponde con $(x, y) = (0, 1)$, luego $\mathrm{N}(\bar{\epsilon}) = -1$, por lo que $\bar{\epsilon}$ es una unidad, al igual que ϵ. De hecho, se comprueba inmediatamente que ϵ es la unidad fundamental de $\mathbb{Z}\big[\sqrt{85}\,\big]$. Para calcularla necesitaríamos el desarrollo en fracción continua

$$-\frac{1+\sqrt{85}}{2} = [4, \bar{9}],$$

pero en realidad basta observar que su primer término es 4, pues en cuanto calculamos el primer convergente $(p_0, q_0) = (4, 1)$ obtenemos $\epsilon = 4 + \omega$, que ya es una unidad, luego es la unidad fundamental.

Por lo tanto, las soluciones de la ecuación se corresponden con los números $\pm\epsilon^n$, donde n recorre los números enteros. Podemos prescindir de los signos negativos, ya que si la solución asociada a ϵ^n es (x_n, y_n), entonces la asociada a $-\epsilon^n$ es $(-x_n, -y_n)$. Claramente $(x_0, y_0) = (1, 0)$ y

$$x_{n+1} + y_{n+1}(-\bar{\epsilon}) = \epsilon^{n+1} = (x_n + y_n(-\bar{\epsilon}))\epsilon = x_n\epsilon + y_n,$$

donde hemos usado que $\epsilon\bar{\epsilon} = \mathrm{N}(\epsilon) = -1$. Usando además que $\epsilon + \bar{\epsilon} = 9$, obtenemos que

$$x_{n+1} + y_{n+1}(-\bar{\epsilon}) = x_n(9 - \bar{\epsilon}) + y_n = 9x_n + y_n + x_n(-\bar{\epsilon}),$$

lo que nos da las relaciones recurrentes

$$(x_{n+1}, y_{n+1}) = (9x_n + y_n, x_n).$$

Por lo tanto, las soluciones de la ecuación correspondientes a exponentes naturales son

$$(1,0), \quad (9,1), \quad (82,9), \qquad (747,82), \quad \ldots$$

Más fácilmente, si definimos $h_{-1} = 0$, $h_0 = 1$, $h_{n+1} = 9h_n + h_{n-1}$, es decir, la sucesión

$$[0], \quad 1, \quad 9, \quad 82, \quad 747, \quad \ldots$$

la solución correspondiente a ϵ^n es (h_n, h_{n-1}).

Ejercicio: Comprobar que las soluciones correspondientes a exponentes negativos son

$$(0,-1), \quad (1,-9), \quad (-9,82), \quad (-82,747), \quad \ldots$$

Obviamente, si queremos únicamente las soluciones de $x^2 - 9xy - y^2 = 1$ o de $x^2 - 9xy - y^2 = -1$ sólo tenemos que quedarnos con los términos pares o impares, respectivamente, de la sucesión de soluciones. ∎

El ejemplo anterior tiene una consecuencia interesante:

Teorema 12.11 *Las ecuaciones $x^3 + y^3 = z^3 \pm 1$ tienen infinitas soluciones enteras con $x, y, z \neq \pm 1$.*

Demostración: Partimos de la identidad

$$(-t^2 + 7t + 9)^3 + (2t^2 + 10)^3 = (2t^2 + 4t + 12)^3 + (-t^2 - 9t + 1)^3$$

que obtuvimos tras el teorema 2.5. Sabemos que para $t = 0$ se reduce a

$$10^3 + 9^3 = 12^3 + 1.$$

El último término sólo vale 1 cuando $t = 0$ o $t = -9$, pero podemos obtener muchas más soluciones enteras a partir de valores racionales del parámetro t. En efecto, si hacemos $t = v/u$ y multiplicamos la ecuación por u^6 obtenemos

$$(9u^2 + 7uv - v^2)^3 + (10u^2 + 2v^2)^3 = (12u^2 + 4uv + 2v^2)^3 + (u^2 - 9uv - v^2)^3.$$

Ahora sólo tenemos que sustituir (u, v) por los valores (x_n, y_n) del ejemplo anterior, con lo que el último cubo se reduce a $(-1)^n$. Así, las primeras soluciones que obtenemos son

$$(x_0, y_0, z_0) = (9, 10, 12), \quad (x_1, y_1, z_1) = (791, 812, 1\,010),$$

$$(x_2, y_2, z_2) = (65\,601, 67\,402, 83\,802).$$

Notemos que las soluciones obtenidas de este modo tienen sus tres componentes positivas, pues las soluciones (u, v) dadas por el ejemplo anterior (para índices positivos) cumplen $u > v$, y es claro entonces que al sustituir en las tres formas cuadráticas obtenemos números positivos.

Las soluciones obtenidas de este modo no son ni mucho menos las únicas. Otros ejemplos de soluciones que pueden obtenerse fácilmente mediante una búsqueda son

$$6^3 + 8^3 = 9^3 - 1, \qquad 64^3 + 94^3 = 103^3 + 1, \qquad 73^3 + 144^3 = 150^3 + 1, \quad \ldots$$

∎

1

$$M = \left\langle 1, \frac{11 + 3\sqrt{15}}{2} \right\rangle, \quad \text{o} \quad M' = \left\langle 2, 11 + 3\sqrt{15} \right\rangle$$

y tratamos de obtener a partir de él una forma cuadrática como

$$\mathrm{N}\left(x + y\frac{11+3\sqrt{15}}{2}\right) = x^2 + 11xy - \frac{7}{2}y^2,$$

$$\mathrm{N}(2x + y(11+3\sqrt{15})) = 4x^2 + 44x - 14y^2,$$

vemos que en el primer caso no nos queda con coeficientes enteros, y hay que multiplicarla por 2, mientras que en el segundo caso no nos queda primitiva, y hay que dividirla entre 2. En general puede quedar cualquier cosa. Por ejemplo:

$$\mathrm{N}\left(\left(\frac{3+2\sqrt{7}}{5}\right)x + \left(\frac{1-6\sqrt{7}}{3}\right)y\right) = -\frac{19x^2}{25} + \frac{58xy}{5} - \frac{251y^2}{9}.$$

Para que esta forma pase a tener coeficientes enteros primos entre sí hay que multiplicarla por 225. Sucede que no hace falta esperar a ver qué sale para saber cuál tiene que ser ese factor corrector para acabar con una forma primitiva, sino que éste depende únicamente del módulo considerado. Vamos a ver que es así.

Definición 12.12 Sea $M = \langle \alpha, \beta \rangle$ un módulo completo de un cuerpo cuadrático k, sea $\mathcal{O}_M = \langle 1, \gamma \rangle$ su anillo de coeficientes, de modo que

$$(\alpha, \beta) = (1, \gamma)A,$$

para cierta matriz A con coeficientes racionales. Sabemos que

$$\Delta[M] = \Delta[\alpha, \beta] = (\det A)^2 \Delta[1, \gamma] = \det(A)^2 \Delta[\mathcal{O}_M].$$

Definimos la *norma* de M como

$$\mathrm{N}(M) = |\det A| = \sqrt{\frac{\Delta[M]}{\Delta[\mathcal{O}_M]}}.$$

Así $\mathrm{N}(M)$ es un número racional positivo caracterizado por la relación

$$\Delta[M] = \mathrm{N}(M)^2 \Delta[\mathcal{O}_M].$$

Vamos a probar que, si $M = \langle \alpha, \beta \rangle$, entonces $\mathrm{N}(M)$ es justo el valor por el que tenemos que dividir la forma cuadrática $\mathrm{N}(x\alpha + y\beta)$ para que tenga coeficientes enteros y sea primitiva. Antes probamos algunos hechos elementales:

1. *Si M es un módulo de un cuerpo cuadrático k y γ es un elemento no nulo de k, entonces*

   $$\mathrm{N}(\gamma M) = |\mathrm{N}(\gamma)|\,\mathrm{N}(M).$$

   En efecto, como M y γM tienen el mismo anillo de coeficientes $\mathcal{O}_M$, de la relación

   $$\Delta[\gamma M] = \mathrm{N}(\gamma)^2 \Delta[M] = \mathrm{N}(\gamma)^2\,\mathrm{N}(M)^2 \Delta[\mathcal{O}_M],$$

   se desprende inmediatamente la relación del enunciado.

2. *Si $\mathcal{O}_m$ es un orden de un cuerpo cuadrático, entonces* $\mathrm{N}(\mathcal{O}_m) = 1$.
   Esto es consecuencia de que $\mathcal{O}_m$ es su propio anillo de coeficientes.

3. *En las condiciones del teorema* 12.9 *se cumple que* $\mathrm{N}(M) = 1/a$.
   En efecto,
   $$(1, \gamma) = (1, a\gamma) \begin{pmatrix} 1 & 0 \\ 0 & 1/a \end{pmatrix}$$
   y $\mathrm{N}(M)$ es el determinante de la matriz.

Nota Tal vez el lector considere —con razón— que la definición que hemos dado de norma de un módulo es un tanto artificiosa. Cuanto menos, es artificioso que a este concepto lo llamemos precisamente "norma". Sin embargo, la primera de las propiedades precedentes muestra una conexión entre este concepto de norma de un módulo de un cuerpo cuadrático k y la norma que ya teníamos definida en k. Esta relación se verá más claramente en el capítulo próximo. ∎

Teorema 12.13 *Para cada base (α, β) de un módulo completo M de un cuerpo cuadrático $k = \mathbb{Q}(\sqrt{d})$, la forma cuadrática*

$$f(x, y) = \frac{\mathrm{N}(x\alpha + y\beta)}{\mathrm{N}(M)} = ax^2 + bxy + cy^2. \tag{12.1}$$

tiene coeficientes enteros, es primitiva, tiene el mismo discriminante que el anillo de coeficientes de M y es definida positiva cuando dicho discriminante es negativo.

Demostración: Llamemos $\gamma = -\beta/\alpha$, que es un elemento de k. Entonces γ es raíz de un único polinomio $Ax^2 + Bx + C$ con coeficientes enteros racionales tales que $(A, B, C) = 1$ y $A > 0$. Así

$$A\left(x + \frac{\beta}{\alpha}\right)\left(x + \frac{\bar{\beta}}{\bar{\alpha}}\right) = Ax^2 + Bx + C.$$

Multiplicando y dividiendo por $\alpha\bar{\alpha}$:

$$\frac{A}{\mathrm{N}(\alpha)}(x\alpha + \beta)(x\bar{\alpha} + \bar{\beta}) = Ax^2 + Bx + C,$$

de donde

$$\mathrm{N}(x\alpha + y\beta) = \frac{\mathrm{N}(\alpha)}{A}(Ax^2 + Bxy + Cy^2).$$

Por las observaciones previas al teorema, la norma de $M = \langle \alpha, \beta \rangle = \alpha \langle 1, \gamma \rangle$ es $|\mathrm{N}(\alpha)|/A$, y por consiguiente

$$ax^2 + bxy + cy^2 = \mathrm{N}(\alpha x + \beta y)/\mathrm{N}(M) = \pm(Ax^2 + Bxy + Cy^2),$$

luego la forma que le hemos asignado a la base (α, β) según (12.1) es primitiva.

Más aún, el anillo de coeficientes de $M = \langle \alpha, \beta \rangle$ es el mismo que el de $\langle 1, \gamma \rangle$, es decir, el orden $\langle 1, A\gamma \rangle$, y, por 12.9, su discriminante es $B^2 - 4AC = b^2 - 4ac$, es decir, el discriminante de la forma asociada.

Además $a = \mathrm{N}(\alpha)/\mathrm{N}(M)$, y por lo tanto si el discriminante es negativo, o sea, si el cuerpo es imaginario, entonces $a > 0$, luego la forma es definida positiva. ∎

En la práctica, a la hora de calcular la norma de un módulo, en lugar de recurrir a la definición, es más práctico calcular la forma cuadrática asociada a una cualquiera de sus bases y ver cuál es el factor (positivo) necesario para que hacerla primitiva. Por ejemplo, en vista del teorema anterior y del cálculo previo a la definición 12.12, no necesitamos ninguna comprobación adicional para concluir que

$$\mathrm{N}\left(\left\langle \frac{3+2\sqrt{7}}{5}, \frac{1-6\sqrt{7}}{3} \right\rangle\right) = \frac{1}{225}.$$

Nota Casi es inmediato que toda forma cuadrática irreducible primitiva (definida positiva si su discriminante es negativo) puede expresarse en términos de un módulo M según el teorema 12.13. El ejemplo siguiente ilustra el único detalle que nos falta tener en cuenta para poder afirmar que esto es así:

Consideremos la forma cuadrática $f(x, y) = -x^2 + 4xy + 2y^2$, de discriminante $D = 24$. Las raíces del polinomio $-x^2 + 4x + 2$ son $2 \pm \sqrt{6}$, lo que nos lleva a que

$$f(x, y) = -(x - 2 + \sqrt{6})(x - 2 - \sqrt{6}).$$

Sin embargo, no podemos tomar el módulo $M_0 = \langle 1, -2 + \sqrt{6} \rangle$, pues entonces

$$\mathrm{N}(x + y(-2 + \sqrt{6})) = x^2 - 4xy - 2y^2$$

luego $\mathrm{N}(M) = 1$ y el teorema 12.13 no nos da la forma $f(x, y)$, sino $-f(x, y)$. Éste es el único problema que podemos encontrarnos al buscar un módulo que determine una forma dada factorizando la forma, que el módulo que obtengamos no nos proporcione f, sino $-f$. Notemos que esto sólo puede suceder si f tiene discriminante positivo, ya que si es negativo, tanto f como la forma proporcionada por M tienen que ser definidas positivas, luego no pueden tener signos opuestos.

Ahora bien, en general, si un módulo $M = \langle \alpha, \beta \rangle$ nos da una forma f de discriminante positivo y γ es cualquier elemento de norma negativa (por ejemplo $\gamma = \sqrt{D}$), entonces el módulo $\gamma M = \langle \gamma\alpha, \gamma\beta \rangle$ tiene asociada la forma

$$\frac{\mathrm{N}(x\gamma\alpha + y\gamma\beta)}{\mathrm{N}(\gamma M)} = \frac{\mathrm{N}(\gamma)\,\mathrm{N}(x\alpha + y\beta)}{|\,\mathrm{N}(\gamma)|\,\mathrm{N}(M)} = -f(x, y).$$

En nuestro ejemplo, podemos tomar

$$M = \sqrt{6}M_0 = \left\langle \sqrt{6}, 6 - 2\sqrt{6} \right\rangle,$$

y así

$$\mathrm{N}(\sqrt{6}+y(6-2\sqrt{6})y) = -6x^2+24xy+12y^2,$$

de donde deducimos que $\mathrm{N}(M)=6$, luego

$$\frac{\mathrm{N}(\sqrt{6}+y(6-2\sqrt{6})y)}{\mathrm{N}(M)} = -x^2+4xy+2y^2.$$

■

El teorema siguiente expresa en general lo que acabamos de observar:

Teorema 12.14 *Si $f(x,y)$ es una forma cuadrática irreducible primitiva de discriminante D (definida positiva si $D<0$), existe un módulo completo $M=\langle\alpha,\beta\rangle$ tal que*

$$f(x,y) = \frac{\mathrm{N}(x\alpha+y\beta)}{\mathrm{N}(M)}.$$

Demostración: Si $f(x,y)=ax^2+bxy+cy^2$ y α es una raíz del polinomio ax^2+bx+c, sabemos que

$$f(x,y) = a(x-y\alpha)(x-y\bar{\alpha}) = aN(x-y\alpha),$$

luego si llamamos $M=\langle 1,-\alpha\rangle$, tenemos que $\mathrm{N}(M)=1/|a|$ y

$$\frac{\mathrm{N}(x-y\alpha)}{\mathrm{N}(M)} = \pm f.$$

Si $D<0$ el signo tiene que ser positivo, pues la forma del miembro izquierdo es definida positiva. Si $D>0$ y el signo es negativo, tomamos un elemento γ de $\mathbb{Q}\big(\sqrt{D}\,\big)$ de norma negativa, y entonces, según hemos razonado en general en la nota precedente, la base $\gamma, -\gamma\alpha$ del módulo γM se corresponde con la forma f.

■

Módulos enteros De entre todos los módulos, hay unos que son especialmente cómodos de manejar y veremos que siempre podemos restringirnos a trabajar con ellos sin pérdida de generalidad:

Definición 12.15 Un módulo de un cuerpo cuadrático es *entero* si está contenido en su anillo de coeficientes. En caso contrario se dice que es *fraccional.*

Si un módulo $M=\langle\alpha,\beta\rangle$ es entero y su anillo de coeficientes es $\mathcal{O}=\langle 1,\gamma\rangle$, entonces las coordenadas de α y β en la base $1,\gamma$ son enteras, luego el determinante de la matriz formada por ellas es un número entero y concluimos que $\mathrm{N}(M)$ es un número natural.

Veamos algunas propiedades de los módulos enteros:

Teorema 12.16 *Sea $k=\mathbb{Q}\big(\sqrt{d}\,\big)$ un cuerpo cuadrático y sea $\omega=\sqrt{d}$ o bien $\omega=(1+\sqrt{d})/2$ según el resto de d módulo 4. Entonces:*

1. *Si M es un módulo completo de la forma $M = k\langle a, b+m\omega\rangle \subset \mathcal{O}_m$, con a, b, k enteros racionales, entonces los elementos de $\mathcal{O}_m$ son coeficientes de M si y sólo si $a \mid \mathrm{N}(b+m\omega)$.*

2. *Todo módulo $M \subset \mathcal{O}_m$ tal que los elementos de $\mathcal{O}_m$ sean coeficientes de M puede expresarse de esta forma.*

3. *El anillo de coeficientes de M es exactamente el orden $\mathcal{O}_m$ si y sólo si $(a, b, m, \mathrm{N}(b+m\omega)/a) = 1$, y en tal caso $\mathrm{N}(M) = k^2|a|$.*

DEMOSTRACIÓN: Sea M un módulo completo contenido en $\mathcal{O}_m$. Sea k el mayor número natural que divida (en $\mathcal{O}_m$) a todos los elementos de a. En particular $M = \langle k\alpha, k\beta\rangle$, con α, $\beta \in \mathcal{O}_m$, luego, llamando $M' = \langle \alpha, \beta\rangle$, tenemos que $M = kM'$ y M' es un módulo completo contenido en $\mathcal{O}_m$ tal que no hay ningún natural mayor que 1 que divida a todos sus elementos. Obviamente los elementos de $\mathcal{O}_m$ serán coeficientes de M si y sólo si lo son de M'.

Observemos ahora que M' admite una base de la forma $M' = \langle a, b+cm\omega\rangle$. Esto lo hemos visto realmente en la prueba del teorema 12.2, pero es fácil aislar el argumento: en principio, M' admite una base de la forma $u+vm\omega, b+cm\omega$, pero sabemos que podemos sustituir uno de sus elementos por la suma o la resta de ambos y, si ambos generadores tienen el coeficiente de $m\omega$ no nulo, podemos hacerlo de modo que el nuevo generador tenga un coeficiente de $m\omega$ de valor absoluto menor que el precedente. Tras un número finito de pasos tenemos que llegar a una base en la que uno de los generadores tenga coeficiente de $m\omega$ nulo.

Es claro que para que los elementos de $\mathcal{O}_m = \langle 1, m\omega\rangle$ sean coeficientes de M' basta que lo sea $m\omega$, y esto equivale a su vez a que $am\omega$ y $m\omega(b+cm\omega)$ estén en M'.

Que am esté en M' equivale a que existan enteros u y v de manera que $am\omega = ua + v(b+cm\omega)$. Necesariamente, $c \neq 0$ y $v = a/c$ y $u = -b/c$. Así pues, $am\omega$ está en M' si y sólo si $c \mid a$ y $c \mid b$. Pero si esto sucede, entonces c divide a todos los elementos de M', luego $c = 1$. En definitiva, $am\omega$ está en M' si y sólo si $c = 1$.

Admitiendo $c = 1$, vamos a ver cuándo se cumple que $m\omega(b+m\omega)$ está en M'. Esto equivale a

$$(m(\omega+\bar{\omega}) - m\bar{\omega})(b+m\omega) \in M'.$$

Sumando y restando b queda

$$(-b-m\bar{\omega})(b+m\omega) + (m(\omega+\bar{\omega})+b)(b+m\omega) \in M'.$$

Ahora bien, $\omega+\bar{\omega} = 0, 1$, luego el segundo sumando es un múltiplo entero de un elemento de M', luego está seguro en M', y la condición equivale a que

$$(-b-m\bar{\omega})(b+m\omega) = -\mathrm{N}(b+m\omega) \in M',$$

o también a que $\mathrm{N}(b+m\omega) = ua + v(b+m\omega)$, lo que equivale a su vez a que $a \mid \mathrm{N}(b+m\omega)$ (pues si se cumple la primera igualdad, necesariamente $v = 0$).

Con esto hemos demostrado las dos primeras afirmaciones del enunciado. Para la última supongamos que un primo p cumple $a = pa'$, $b = pb'$, $m = pm'$, $ap \mid \mathrm{N}(b + m\omega) = p^2\,\mathrm{N}(b' + m'\omega)$. Entonces $M = kp\,\langle a' + b' + m'\omega\rangle$ de modo que $a' \mid \mathrm{N}(b' + m'\omega)$, luego, por la parte ya probada, los elementos de $\mathcal{O}_{m'}$ serían también coeficientes de M.

Recíprocamente, si el anillo de coeficientes de M es $\mathcal{O}_{m'}$ con $\mathcal{O}_m \subsetneq \mathcal{O}_{m'}$, entonces $m' \mid m$ y podemos tomar un primo p que divida a m/m'. Así $m = pm''$, donde $m' \mid m''$, luego $\mathcal{O}_{m''} \subset \mathcal{O}_{m'}$. Por la parte ya probada, p tiene que dividir a todos los elementos de M (en $\mathcal{O}_{m''}$), ya que, al eliminar el mayor divisor común de todos los elementos de M en $\mathcal{O}_{m''}$ el coeficiente de ω tiene que ser exactamente m''. Equivalentemente, $a = pa'$, $b = pb'$ y $M = kp\,\langle a', b' + m''\omega\rangle$, y además, también por la parte ya probada, $a' \mid \mathrm{N}(b' + m''\omega)$, luego $pa = p^2a' \mid \mathrm{N}(a, b + m\omega)$, luego $p \mid \mathrm{N}(a, b + m\omega)/a$.

Finalmente, si se da este caso, aplicamos la definición de norma 12.12, donde la matriz A es, en este caso,

$$A = \begin{pmatrix} ka & kb \\ 0 & k \end{pmatrix}$$

■

12.2 La correspondencia entre módulos y formas

El teorema 12.13 hace corresponder una forma cuadrática a cada base de un módulo, pero ¿qué sucede si consideramos otra base del mismo módulo?

Cambios de base Si α, β y γ, δ son dos bases de un cuerpo cuadrático k, entonces

$$\gamma = r\alpha + s\beta, \qquad \delta = r'\alpha + s'\beta,$$

para ciertos números enteros r, s, r', s' unívocamente determinados. Matricialmente:

$$\begin{pmatrix} \gamma \\ \delta \end{pmatrix} = \begin{pmatrix} r & s \\ r' & s' \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

La matriz que aparece en esta ecuación se llama *matriz de cambio de base.* Más precisamente, es la matriz de cambio de la base (γ, δ) a la base (α, β), porque sirve para transformar las coordenadas de un número ξ en la primera base en las coordenadas en la segunda. En efecto, si las coordenadas son (x, y), esto significa que

$$\xi = x\gamma + y\delta = x(r\alpha + s\beta) + y(r'\alpha + s'\beta) = (xr + yr')\alpha + (xs + ys')\beta,$$

luego las coordenadas de ξ en la base (α, β) son

$$(x', y') = (x, y) \begin{pmatrix} r & s \\ r' & s' \end{pmatrix}.$$

Recíprocamente, si una matriz A determina de este modo la correspondencia entre las coordenadas de ambas bases, tiene que ser necesariamente la matriz de cambio de base. Basta aplicar esta relación a $(x, y) = (1, 0)$ (que son las coordenadas de γ en la base γ, δ) para concluir que (r, s) son las coordenadas de γ en la base α, β y a $(x, y) = (0, 1)$ para concluir que (r', s') son las coordenadas de δ en la base α, β.

Si llamamos A a la matriz de cambio de base de (γ, δ) a (α, β) y B a la matriz de cambio de base de (α, β) a (γ, δ), tenemos que

$$(x', y') = A(x, y) = AB(x', y'),$$

luego AB es la matriz de cambio de base de α, β a α, β, pero esta matriz es obviamente I, luego por la unicidad $AB = I$.

En otras palabras, las matrices de cambio de base en ambos sentidos entre dos bases dadas son mutuamente inversas.

Si ambas bases lo son de un mismo módulo M, entonces tanto A como $B = A^{-1}$ tienen coeficientes enteros, pues las coordenadas de los elementos de M en una base de M son enteras, luego la relación $|A||B| = 1$ implica que $|A| = \pm 1$.

Recíprocamente, es claro que cada matriz con coeficientes racionales de determinante no nulo determina una base de k a partir de otra dada y que cada matriz con coeficientes enteros de determinante ± 1 determina una base de un módulo M a partir de otra dada.

Ejemplo Consideremos el módulo $M = \langle 77, 38 + 3\sqrt{15}\rangle$. La forma cuadrática asociada a esta base es $77x^2 + 76xy + 17y^2$. Tras el teorema 11.23 vimos que es estrictamente equivalente a la forma reducida $2x^2 - 22xy - 7y^2$ a través del cambio de variables

$$(x, y) = (x', y') \begin{pmatrix} 1 & -3 \\ 12 & -35 \end{pmatrix}.$$

Según acabamos de ver, una base de M que determina esta forma cuadrática es

$$\begin{pmatrix} 1 & -3 \\ 12 & -35 \end{pmatrix} \begin{pmatrix} 77 \\ 38 + 3\sqrt{15} \end{pmatrix} = \begin{pmatrix} -37 - 9\sqrt{15} \\ -406 - 105\sqrt{15} \end{pmatrix},$$

es decir, $M = \langle -37 - 9\sqrt{15}, -406 - 105\sqrt{15}\rangle = \langle 37 + 9\sqrt{15}, 406 + 105\sqrt{15}\rangle$, y el lector puede comprobar que la forma cuadrática asociada a cualquiera de estas dos bases es la requerida. ∎

Similitud de módulos Ahora es claro que si f es la forma cuadrática asociada a la base α, β, entonces la asociada a γ, δ es

$$f'(x, y) = \frac{\mathrm{N}((r\alpha + s\beta)x + (r'\alpha + s'\beta)y)}{\mathrm{N}(M)} = \frac{\mathrm{N}((rx + r'y)\alpha + (sx + s'y)\beta)}{\mathrm{N}(M)}$$
$$= f(rx + r'y, sx + s'y),$$

luego se trata de la forma cuadrática que resulta de aplicar a $f(x', y')$ el cambio de variables

$$(x', y') = (x, y) \begin{pmatrix} r & s \\ r' & s' \end{pmatrix},$$

que tiene determinante ± 1. Así hemos demostrado lo siguiente:

> *La aplicación $(\alpha, \beta) \mapsto f(x, y)$ definida en el teorema* 12.13 *hace corresponder las bases de un módulo completo M fijo con todas las formas cuadráticas de una misma clase de equivalencia. Las formas asociadas a dos bases de M están relacionadas por el cambio de base cuya matriz coincide con la la matriz de cambio de base.*

En otras palabras, a cada módulo completo M le hemos asociado, no una forma cuadrática, sino una clase de equivalencia completa de formas cuadráticas. Pero no es cierto que módulos distintos tengan asociadas clases de equivalencia distintas (sería imposible, porque hay infinitos módulos y sólo una cantidad finita de clases de equivalencia). A este respecto observamos lo siguiente:

Si γ es un elemento no nulo de k y α, β es una base de M cuya forma asociada es $f(x, y)$, entonces $\gamma\alpha, \gamma\beta$ es una base de γM cuya forma asociada es

$$\frac{\mathrm{N}(\gamma\alpha x + \gamma\alpha y)}{\mathrm{N}(\gamma M)} = \frac{\mathrm{N}(\gamma)\,\mathrm{N}(\alpha x + \beta y)}{|\,\mathrm{N}(\gamma)|\,\mathrm{N}(M)} = \pm f(x, y),$$

donde el signo es el de $\mathrm{N}(\gamma)$.

Esto nos lleva a las definiciones siguientes:

Definición 12.17 Si M y M' son dos módulos completos de un mismo cuerpo cuadrático k, diremos que son *(estrictamente) similares* si existe un γ en k no nulo (con $\mathrm{N}(\gamma) > 0$) tal que $M' = \gamma M$.

Obviamente, si dos módulos son estrictamente similares, entonces son similares, y, si el cuerpo k es imaginario, entonces no hay elementos de norma negativa, luego también se cumple el recíproco.

He aquí algunas propiedades elementales:

1. **Propiedad reflexiva** *Todo módulo completo es estrictamente similar a sí mismo.*

   Basta tener en cuenta que $M = 1M$.

2. **Propiedad simétrica** *Si un módulo completo M es (estrictamente) similar a otro M', entonces M' es (estrictamente) similar a M.*

   En efecto, si $M' = \gamma M$, entonces $M = \gamma^{-1} M'$ y $\mathrm{N}(\gamma^{-1}) = \mathrm{N}(\gamma)^{-1}$ tiene el mismo signo que $\mathrm{N}(\gamma)$.

3. **Propiedad transitiva** *Si un módulo completo M es (estrictamente) similar a otro M' y M' es (estrictamente) similar a M'', entonces M es (estrictamente) similar a M''.*

   En efecto, si $M' = \gamma M$ y $M'' = \gamma' M'$, entonces $M'' = \gamma'\gamma M$ y si γ y γ' tienen norma positiva, $\mathrm{N}(\gamma'\gamma) = \mathrm{N}(\gamma')\,\mathrm{N}(\gamma) > 0$.

4. *Dos módulos completos similares de un mismo cuerpo cuadrático tienen el mismo anillo de coeficientes.*

   Esto lo hemos demostrado antes del teorema 12.9.

5. *Si dos módulos completos de un mismo cuerpo cuadrático son estrictamente similares, sus clases de equivalencia de formas cuadráticas asociadas son la misma.*

   Esto lo hemos demostrado antes de la definición 12.17.

6. *Todo módulo completo es estrictamente similar a un módulo entero.*

   En efecto, si $M = \langle \alpha, \beta\rangle$, entonces $M = \alpha\,\langle 1, \gamma\rangle$, donde $\gamma = \beta/\alpha$ y, por el teorema 12.9, su anillo de coeficientes es de la forma $\mathcal{O} = \langle 1, a\gamma\rangle$, para cierto entero a. Por lo tanto, M es similar a $\langle 1, \gamma\rangle$, que a su vez es similar a $M' = \langle a, a\gamma\rangle \subset \mathcal{O}$. Si la similitud no es estricta es que estamos en un cuerpo cuadrático real, por lo que $\mathcal{O}$ contiene elementos de norma negativa (basta tomar $\delta = k\sqrt{d}$, para un entero k adecuado). Así, si $M = \xi M'$, con $\mathrm{N}(\xi) < 0$, tenemos que $M = (\xi/\delta)\delta M'$, donde $\delta M' \subset \mathcal{O}$ y $\mathrm{N}(\xi/\delta) > 0$.

Las tres primeras propiedades hacen que podamos hablar de clases de similitud y de similitud estricta de módulos completos, exactamente igual que hablamos de clases de equivalencia y de equivalencia estricta de formas cuadráticas.

La quinta propiedad nos dice que, al igual que al estudiar una forma cuadrática podemos sustituirla por otra equivalente que nos resulte más conveniente, al asociar un módulo a una forma dada podemos sustituirlo por otro estrictamente similar que nos resulte más conveniente, y la sexta propiedad nos da un ejemplo de ventaja que podemos obtener eligiendo un módulo adecuadamente en una clase de similitud estricta: podemos elegirlo entero, lo que nos permite trabajar exclusivamente con enteros algebraicos.

Orientación de bases La quinta de las propiedades que acabamos de probar sobre la similitud de módulos concierne específicamente a la similitud estricta, pues si dos módulos $M' = \gamma M$ son similares, pero $\mathrm{N}(\gamma) < 0$, sabemos que si al primero le corresponde una clase de equivalencia de formas $[f]$, al segundo le corresponde la clase $[-f]$, que no es necesariamente la misma que la primera (más adelante veremos ejemplos concretos).

De este modo, a cada clase de similitud estricta de módulos le podemos asignar una clase de equivalencia no estricta de formas cuadráticas. Sucede que esta correspondencia no es biunívoca, pero se puede refinar para obtener una correspondencia biunívoca $[M] \leftrightarrow [f]$ entre clases de similitud estricta de módulos y clases de equivalencia estricta de formas.

Sin embargo, en una primera aproximación al problema nos encontramos con un inconveniente, y es que, al recorrer las bases de un módulo completo M, las formas asociadas recorren toda una clase de equivalencia de formas cuadráticas, las cuales no tienen por qué ser estrictamente equivalentes, sino que pueden determinar dos clases de equivalencia estrictas distintas. Entonces ¿cuál de las dos le asignamos a la clase de similitud estricta de M? Para resolver este problema necesitamos introducir un último concepto:

Definición 12.18 Diremos que dos bases de un mismo cuerpo cuadrático k *tienen la misma orientación* si la matriz de cambio de base tiene determinante positivo. En caso contrario se dice que tienen *orientaciones opuestas.*

Veamos algunas propiedades elementales:

1. **Propiedad reflexiva** *Toda base de un cuerpo cuadrático tiene la misma orientación que ella misma.*

   Porque la matriz de cambio de base es I, que tiene determinante 1.

2. **Propiedad simétrica** *Si una base (α, β) tiene la misma orientación que otra (γ, δ), entonces (γ, δ) tiene la misma orientación que (α, β).*

   Porque las matrices de cambio de base son mutuamente inversas, luego cumplen $|A||B| = 1$, luego ambas tienen el mismo signo.

3. **Propiedad transitiva** *Si una base tiene la misma orientación que otra y ésta tiene la misma orientación que una tercera, entonces la primera tiene la misma orientación que la tercera.*

   Porque la matriz de cambio de base de la primera a la tercera es el producto de la matriz de cambio de base de la primera a la segunda por la matriz de cambio de base de la segunda a la tercera y, si ambas tienen determinante positivo, lo mismo le sucede al producto.

4. *Las bases (α, β) y (β, α) tienen orientaciones opuestas.*

   Porque la matriz de cambio de base es

   $$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

   que tiene determinante -1.

5. *Las bases (α, β) y $(\bar{\alpha}, \bar{\beta})$ tienen orientaciones opuestas.*

   Pongamos que $\alpha = r + s\sqrt{d}$, $\beta = r' + s'\sqrt{d}$. Entonces

   $$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} r & s \\ r' & s' \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{d} \end{pmatrix}, \qquad \begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \end{pmatrix} = \begin{pmatrix} r & -s \\ r' & -s' \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{d} \end{pmatrix},$$

   luego

   $$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} r & s \\ r' & s' \end{pmatrix} \begin{pmatrix} r & -s \\ r' & -s' \end{pmatrix}^{-1} \begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \end{pmatrix}.$$

   Las dos matrices tienen determinantes opuestos, luego el producto de una por la inversa de la otra tiene determinante -1.

6. *Si* $\mathrm{N}(\gamma) > 0$*, entonces las bases* (α, β) *y* $(\gamma\alpha, \gamma\beta)$ *tienen la misma orientación.*

   Pongamos que $\alpha = r + s\sqrt{d}$, $\beta = r' + s'\sqrt{d}$, $\gamma = u + v\sqrt{d}$. Entonces

$$\begin{aligned}\gamma\alpha &= ru + svd + (rv + su)\sqrt{d},\\ \gamma\beta &= r'u + s'vd + (r'v + s'u)\sqrt{d},\end{aligned}$$

   luego la matriz de cambio de base de $(\gamma\alpha, \gamma\beta)$ a $(1, \sqrt{d})$ es

$$\begin{pmatrix} ru + svd & rv + su \\ r'u + s'vd & r'v + s'u \end{pmatrix} = \begin{pmatrix} r & s \\ r' & s' \end{pmatrix} \begin{pmatrix} u & v \\ vd & u \end{pmatrix}$$

   y al tomar determinantes vemos que los de las dos matrices de cambio de base se diferencian en un factor $u^2 - v^2 d = \mathrm{N}(\gamma) > 0$. Por lo tanto, o ambas bases tienen la misma orientación que $(1, \sqrt{d})$, o bien ambas tienen orientación opuesta a $(1, \sqrt{d})$, y en ambos casos tienen la misma orientación.

Diremos que una base de un cuerpo cuadrático k está *positivamente orientada* si tiene la misma orientación que la base $(1, \sqrt{d})$. En caso contrario diremos que está *negativamente orientada.*

De este modo, si dos bases están positivamente orientadas, la matriz de cambio de base tiene determinante positivo. En particular, si dos bases de un mismo módulo completo están positivamente orientadas, la matriz de cambio de base tiene determinante 1.

En todo módulo completo hay bases positivamente orientadas, pues si una base (α, β) no lo está, nos sirve (β, α).

Ahora es claro que, cuando (α, β) recorre todas las bases positivamente orientadas de un módulo completo M, las formas cuadráticas dadas por el teorema 12.13 recorren una clase de equivalencia estricta de formas cuadráticas. Por lo tanto, tenemos una correspondencia $M \mapsto [f]$ que a cada módulo completo le asigna la clase de equivalencia estricta formada por las formas cuadráticas asociadas a sus bases orientadas.

Más aún, hemos visto que si $\mathrm{N}(\gamma) > 0$, entonces la forma cuadrática asociada a una base positivamente orientada (α, β) es la misma que la asociada a $(\gamma\alpha, \gamma\beta)$, que también está positivamente orientada, luego las formas asociadas a las bases positivamente orientadas de un módulo completo M son las mismas que las asociadas a las bases positivamente orientadas de γM, luego a M y γM les corresponde la misma clase de equivalencia estricta de formas cuadráticas.

Por consiguiente, tenemos una correspondencia $[M] \mapsto [f]$ que a cada clase de similitud estricta de módulos, le asigna la clase de equivalencia estricta de las formas asociadas a las bases positivamente orientadas de cualquiera de los módulos de la clase dada.

Por otra parte, el teorema 12.14 nos dice que toda forma cuadrática irreducible primitiva (y definida positiva si su discriminante es negativo) es de la forma

$$f(x,y) = \frac{\mathrm{N}(x\alpha + y\beta)}{\mathrm{N}(M)},$$

para cierta base (α, β) de cierto módulo completo M. Si la base resulta estar negativamente orientada, podemos cambiarla por $(\bar\alpha, \bar\beta)$, pues

$$\mathrm{N}(x\bar\alpha + y\bar\beta) = (x\bar\alpha + y\bar\beta)\overline{(x\bar\alpha + y\bar\beta)} = (x\bar\alpha + y\bar\beta)(x\alpha + y\beta) = \mathrm{N}(x\alpha + y\beta)$$

luego la base $(\bar\alpha, \bar\beta)$ tiene asociada la misma forma cuadrática y está positivamente orientada.

Por consiguiente, toda clase de equivalencia estricta de formas cuadráticas está asociada a una clase de similitud estricta de módulos. Con esto casi hemos demostrado el teorema siguiente:

Teorema 12.19 *Si $\mathcal{O}$ es un orden de discriminante D en un cuerpo cuadrático k, existe una correspondencia biunívoca $[M] \leftrightarrow [f]$ entre las clases de similitud estricta de módulos completos de un cuerpo cuadrático k cuyo anillo de coeficientes es $\mathcal{O}$ y las clases de equivalencia estricta de formas cuadráticas primitivas de discriminante D (definidas positivas si $D < 0$). Dicha correspondencia asocia a cada clase de similitud estricta la clase de las formas cuadráticas asociadas a las bases positivamente orientadas de cualquiera de sus módulos.*

Demostración: Sólo falta probar que a dos clases de similitud estricta distintas no puede corresponderles la misma clase de equivalencia estricta de formas. En efecto, supongamos que dos módulos M y M' tienen asignada la misma clase de equivalencia estricta de formas. Entonces podemos tomar bases orientadas (α, β) de M y (α', β') de M' cuya forma cuadrática asociada sea la misma, digamos $ax^2 + bxy + cy^2$.

En la prueba del teorema 12.13 hemos visto que $a = \mathrm{N}(\alpha)/\mathrm{N}(M)$, luego $\mathrm{N}(\alpha)$ tiene el mismo signo que el coeficiente de x^2, y lo mismo vale para la otra base, luego $\mathrm{N}(\alpha)$ y $\mathrm{N}(\alpha')$ tienen el mismo signo. Por consiguiente las bases $(1, \beta/\alpha)$ y $(1, \beta'/\alpha')$ están ambas orientadas o ninguna lo está, según el signo de $\mathrm{N}(\alpha)$.

Pero $-\beta/\alpha$ y $-\beta'/\alpha'$ son raíces del polinomio irreducible $ax^2 + bx + c$, luego son iguales o conjugados en k, y no pueden ser conjugados porque la conjugación invierte la orientación, así que son iguales. De aquí se sigue que

$$M = \langle \alpha, \beta\rangle = \alpha\,\langle 1, \beta/\alpha\rangle = \alpha\,\langle 1, \beta'/\alpha'\rangle = (\alpha/\alpha')\,\langle \alpha', \beta'\rangle = (\alpha/\alpha')M',$$

luego M y M' son estrictamente similares. ∎

Ejemplo Consideremos las formas cuadráticas de discriminante $D = 316$. Hay seis clases de similitud estricta, con representantes:

$$\begin{array}{ccc} x^2 + 16xy - 15y^2 & -x^2 + 16xy + 15y^2 & 3x^2 + 14xy - 10y^2 \\ -3x^2 + 14xy + 10y^2 & 10x^2 + 14xy - 3y^2 & -10x^2 + 14xy + 3y^2 \end{array}$$

Sus módulos asociados tienen que tener como anillo de coeficientes el orden del mismo discriminante, $\Delta = 4 \cdot 79$, que es el orden $\mathcal{O}_1$ de $k = \mathbb{Q}\big(\sqrt{79}\big)$, es decir, el anillo de enteros algebraicos de este cuerpo, $\mathbb{Z}\big[\sqrt{79}\big]$.

Para calcular la clase asociada al primero tomamos una raíz del polinomio $x^2 + 16x - 15$, por ejemplo

$$\frac{-16 + \sqrt{316}}{2} = -8 - \sqrt{79},$$

de donde obtenemos que

$$x^2 + 16xy - 15y^2 = (x + y(8 + \sqrt{79}))(x + y(8 - \sqrt{79})),$$

luego su clase de módulos asociada es la de

$$M_1 = \left\langle 1, 8 + \sqrt{79} \right\rangle,$$

donde tenemos en cuenta que la base está positivamente orientada, pues

$$\begin{vmatrix} 1 & 0 \\ 8 & 1 \end{vmatrix} = 1 > 0.$$

Este módulo es el mismo que $M_1 = \langle 1, \sqrt{79} \rangle$, cuya forma asociada es $x^2 - 79y^2$, que es, pues otro representante de la misma clase de similitud estricta de formas.

En el caso de $-x^2 + 16xy + 15y^2$ partimos de la raíz

$$\frac{-16 + \sqrt{316}}{-2} = 8 + \sqrt{79},$$

que nos lleva a la factorización

$$-x^2 + 16xy + 15y^2 = -(x + y(-8 - \sqrt{79}))(x + y(-8 + \sqrt{79})),$$

pero no podemos tomar el módulo $\langle 1, -8 - \sqrt{79} \rangle$ (es fácil ver que este módulo es M_1) porque su forma asociada es $x^2 - 16xy - 15y^2$. Tenemos que multiplicarlo por un elemento de norma negativa, por ejemplo $\sqrt{79}$, con lo que obtenemos el módulo

$$M_2 = \sqrt{79}\left\langle 1, -8 - \sqrt{79} \right\rangle = \left\langle \sqrt{79}, -79 - 8\sqrt{79} \right\rangle.$$

La forma asociada a esta base sí que es la correcta, y además está positivamente orientada, pues

$$\begin{vmatrix} 0 & 1 \\ -79 & -8 \end{vmatrix} = 79 > 0.$$

Equivalentemente, usando el principio de que a un generador podemos sumarle un múltiplo de otro, así como que podemos cambiar el signo a cualquier generador:

$$M_2 = \left\langle -79, \sqrt{79} \right\rangle = \left\langle 79, \sqrt{79} \right\rangle.$$

Las dos bases están positivamente orientadas y la forma asociada a la última es $79x^2 - y^2$.

Consideremos ahora la forma $3x^2 + 14xy - 10y^2$. Partimos de la raíz

$$\frac{-14 - \sqrt{316}}{6} = \frac{-7 - \sqrt{79}}{3},$$

que nos da la descomposición

$$3x^2 + 14xy - 10y^2 = 3\left(x + y\frac{7 + \sqrt{79}}{3}\right)\left(x + y\frac{7 - \sqrt{79}}{3}\right),$$

que a su vez nos lleva al módulo

$$M = \left\langle 1, \frac{7 + \sqrt{79}}{3} \right\rangle$$

cuya base está positivamente orientada. No obstante, podemos considerar un módulo más simple estrictamente similar:

$$M_3 = 3M = \left\langle 3, 7 + \sqrt{79} \right\rangle = \left\langle 3, 1 + \sqrt{79} \right\rangle.$$

La forma asociada a la última base es $3x^2 + 2xy - 26y^2$. Igualmente podríamos calcular módulos para las clases de formas restantes, pero podemos seguir un camino más sencillo.

En general, si a un módulo completo $M = \langle \alpha, \beta \rangle$ le corresponde una forma cuadrática $ax^2 + bxy + cy^2$, a la base conjugada $\bar\alpha, \bar\beta$ le corresponde la misma forma, pero si la base dada está positivamente orientada, la conjugada no lo está, luego tenemos que invertir el orden y así, al módulo conjugado $\bar M = \langle \bar\beta, \bar\alpha \rangle$ le corresponde la forma $cx^2 + bxy + ay^2$ (o, si lo preferimos, a través del cambio $x = y'$, $y = -x'$, la forma $ax^2 - bxy + cy^2$).

En nuestro ejemplo, esto nos permite completar la cuarta fila de la tabla siguiente, donde en la segunda columna está el módulo conjugado $M_4 = \bar M_3$ con la base conjugada de la primera base de M_3 en orden inverso para hacerla positiva y con la que resulta de la segunda base manteniendo el orden pero cambiando un signo.

Por otra parte, si $[M]$ es una clase de módulos, podemos considerar la clase $-[M] = [\gamma M]$, donde γ es cualquier elemento de norma negativa (la clase de similitud estricta será la misma en cualquier caso). En la prueba del teorema 12.14 (en la nota precedente, de hecho) hemos visto que si una base de M está asociado a una forma f, entonces la base correspondiente de γM está asociada a $-f$, pero es una base negativamente orientada, luego, cambiando un signo, obtenemos que si una clase $[M]$ está asociada a una clase de formas $[ax^2 + bxy + cy^2]$, entonces $-[M]$ está asociada a la clase $[-ax^2 + bxy - cy^2]$.

En nuestro caso, podemos tomar $M_5 = \sqrt{79}M_3 = \langle -3\sqrt{79}, 79 + 7\sqrt{79} \rangle$, con lo que $-[M_3] = [M_5]$. Equivalentemente,

$$M_5 = \left\langle -3\sqrt{79}, 79 + \sqrt{79} \right\rangle = \left\langle 3 \cdot 79, 79 + \sqrt{79} \right\rangle.$$

Esto nos da la quinta fila de la tabla siguiente. Por último, la clase $-[\bar{M}]$ se corresponde con la clase de formas $[-ax^2 - bxy - cy^2] = [-cx^2 + bxy - ay^2]$, lo que nos da la sexta fila:

$[x^2 + 16xy - 15y^2]$	$[\langle 1, \sqrt{79}\rangle]$	$[x^2 - 79y^2]$
$[-x^2 + 16xy + 15y^2]$	$[\langle 79, \sqrt{79}\rangle]$	$[79x^2 - y^2]$
$[3x^2 + 14xy - 10y^2]$	$[\langle 3, 1 + \sqrt{79}\rangle]$	$[3x^2 + 2xy - 26y^2]$
$[-10x^2 + 14xy + 3y^2]$	$[\langle 3, -1 + \sqrt{79}\rangle]$	$[3x^2 - 2xy - 26y^2]$
$[-3x^2 + 14xy + 10y^2]$	$[\langle -3\sqrt{79}, 79 + \sqrt{79}\rangle]$	$[-3x^2 + 2xy + 26y^2]$
$[10x^2 + 14xy - 3y^2]$	$[\langle 3\sqrt{79}, -79 + \sqrt{79}\rangle]$	$[-3x^2 - 2xy + 26y^2]$

■

Nota En general, hemos probado la correspondencia:

$$\begin{aligned} [M] &\longleftrightarrow [ax^2 + bxy + cy^2] \\ [\overline{M}] &\longleftrightarrow [ax^2 - bxy + cy^2] \\ -[M] &\longleftrightarrow [-ax^2 + bxy - cy^2] \\ -[\overline{M}] &\longleftrightarrow [-ax^2 - bxy - cy^2] \end{aligned}$$

pero no es necesariamente cierto que las cuatro clases de módulos o formas sean distintas dos a dos. De hecho, en el ejemplo anterior sucede que $[M_1] = [\bar{M}_1]$ y $[M_2] = [\bar{M}_2]$, porque se corresponden con formas con $b = 0$, y es claro que cada forma con $b = 0$ es estrictamente equivalente a su "forma conjugada" (la que resulta de cambiarle el signo a b).

Por otro lado, si consideramos clases de similitud y equivalencia no estrictas, es claro que $[M] = -[M]$ (luego $[\bar{M}] = -[\bar{M}]$) y, por otro lado,

$$[ax^2 + bxy + cy^2] = [ax^2 - bxy + cy^2], \quad [-ax^2 + bxy - cy^2] = [-ax^2 - bxy - cy^2],$$

pues una clase se transforma en la otra mediante el cambio de variables $x = -x'$, $y = y'$. Esto se traduce en que la correspondencia entre clases estrictas de módulos y formas no induce una correspondencia análoga entre clases no estrictas. En el ejemplo anterior tenemos tres clases de cada tipo:

$[\langle 1, \sqrt{79}\rangle]$	$[x^2 - 79y^2]$
$[\langle 79, \sqrt{79}\rangle]$	$[79x^2 - y^2]$
$[\langle 3, 1 + \sqrt{79}\rangle]$	$[3x^2 + 2xy - 26y^2]$
$[\langle -3\sqrt{79}, 79 + \sqrt{79}\rangle]$	$[3x^2 - 2xy - 26y^2]$
$[\langle 3, -1 + \sqrt{79}\rangle]$	$[-3x^2 + 2xy + 26y^2]$
$[\langle 3\sqrt{79}, -79 + \sqrt{79}\rangle]$	$[-3x^2 - 2xy + 26y^2]$

Las tres son distintas entre sí, pues si, por ejemplo, $M_3 = \langle 3, 1+\sqrt{79}\rangle$ fuera similar a $M_4 = \langle 3, -1+\sqrt{79}\rangle$, como sabemos que no son estrictamente similares, tendría que ser $M_4 = \gamma M_3$, con $\mathrm{N}(\gamma) < 0$, pero entonces M_4 sería estrictamente similar a $M_5 = \sqrt{79}M_3$, pero sabemos que no es así, pues son representantes de dos clases de similitud estricta distintas. Igualmente se descartan las demás posibilidades, y análogamente se concluye que las tres clases de formas son distintas entre sí.

Y ahora es evidente que no podemos establecer una correspondencia entre las clases no estrictas relajando la correspondencia entre las clases estrictas. Por ejemplo, la imagen de la clase $[\langle 3, 1+\sqrt{79}\rangle] = [\langle -3\sqrt{79}, 79+\sqrt{79}\rangle]$ debería ser $[3x^2 + 2xy - 26y^2]$ si consideramos el primer representante, pero debería ser $[-3x^2 + 2xy + 26y^2]$ si consideramos el segundo, y son dos clases distintas.

Por consiguiente, trabajar con relaciones estrictas es imprescindible para establecer la correspondencia entre clases de módulos y de formas cuadráticas. ∎

Ejercicio: Explicitar la correspondencia entre clases estrictas de módulos y formas cuadráticas de discriminante $D = 328$ y $D = -108$.

Más adelante tendrá interés el hecho siguiente:

Teorema 12.20 *Todo módulo completo cuyo anillo de coeficientes es un orden cuadrático de discriminante D es similar a un módulo entero de norma menor o igual que $\sqrt{|D|/3}$ (y si $D > 0$, incluso de norma menor o igual que $\sqrt{D/4}$).*

Demostración: Sea $M_0 = \langle \alpha, \beta\rangle$ un módulo completo cuyo anillo de coeficientes sea el orden $\mathcal{O}$ de discriminante D. El teorema 12.13 nos da una forma cuadrática primitiva $f(x, y) = ax^2 + bxy + cy^2$ de discriminante D. Cambiando la base dada por otra, podemos suponer que la forma es reducida. Si α es una raíz del polinomio $ax^2 + bx + c$, el módulo $M_1 = \langle 1, -\alpha\rangle$ tiene asociada la forma $\pm f$ (véase la prueba del teorema 12.14). Notemos que la base se puede tomar orientada cambiando α por su conjugado si es preciso. Teniendo en cuenta que también $M = \langle 1, \alpha\rangle$, sabemos (propiedad 3 de la página 422) que $\mathrm{N}(M) = 1/|a|$, luego el módulo $M = aM_1 = \langle a, a\alpha\rangle$ es estrictamente similar a M_1 (luego también está asociado a la clase $[\pm f]$), cumple que $\mathrm{N}(M) = a^2\,\mathrm{N}(M_1) = |a|$ y es un módulo entero, (pues, por 12.9, su anillo de coeficientes es $\langle 1, a\alpha\rangle$).

El signo negativo (en $\pm f$) sólo se da si $a < 0$, en cuyo caso $D > 0$ (pues en caso contrario la forma cuadrática es definida positiva) y, tomando un γ de norma negativa, tenemos que el módulo γM esta asociado a la clase $[f]$, luego es estrictamente similar a M_0. Por lo tanto, en cualquier caso el módulo M es similar a M_0 y tiene norma $|a|$. Cambiando f por la forma estrictamente equivalente $f(y, -x)$ podemos obtener igualmente un módulo similar a M_0 de norma $|c|$.

Ahora, si $D < 0$, el hecho de que la forma f sea reducida implica que $a \le \sqrt{|D|/3}$ (véase la observación tras el teorema 11.19), mientras que si $D > 0$ y $N = \min\{|a|, |c|\}$ tenemos que $N^2 \le |ac| = -ac = (D - b^2)/4 \le D/4$, luego $\mathrm{N}(M) \le \sqrt{D/4}$. ∎

Las formas principales Si D es un número entero tal que $D \equiv 0, 1 \pmod 4$, existen ciertamente formas cuadráticas de discriminante D. La más simple es la que Gauss llamó *forma principal* de discriminante D, que es

$$x^2 - \frac{D}{4}y^2 \qquad \text{o} \qquad x^2 + xy + \frac{1-D}{4}y^2,$$

según el resto de D módulo 4. La clase de equivalencia estricta de la forma principal se llama también *clase principal* de discriminante D y a su vez, su clase de similitud estricta de módulos se llama también *clase principal*. Por extensión, también se llama *clase principal* a la clase de similitud no estricta de módulos que contiene a la clase principal estricta.

Teorema 12.21 *La clase principal (estricta o no) de módulos cuyo anillo de coeficientes es un orden cuadrático $\mathcal{O}$ es $[\mathcal{O}]$.*

DEMOSTRACIÓN: Si $D \equiv 1 \pmod 4$, entonces $D = m^2 d$, con $d \equiv 1 \pmod 4$ y m impar. Una raíz del polinomio $x^2 + x + \frac{1-D}{4}$ es

$$\frac{-1 - m\sqrt{d}}{2},$$

por lo que

$$x^2 + xy + \frac{1-D}{4}y^2 = \mathrm{N}(x + y\frac{1+m\sqrt{d}}{2})$$

y un módulo asociado a la forma principal es (notemos que las bases están orientadas)

$$\left\langle 1, \frac{1+m\sqrt{d}}{2} \right\rangle = \left\langle 1, \frac{m-1}{2} + \frac{1+m\sqrt{d}}{2} \right\rangle = \left\langle 1, m\frac{1+\sqrt{d}}{2} \right\rangle = \mathcal{O}_m.$$

Si $D \equiv 0 \pmod 4$, entonces $D = (2m)^2 d$ y

$$x^2 - \frac{D}{4}y^2 = \mathrm{N}(x + ym\sqrt{d}),$$

luego un módulo asociado a la forma principal es $\left\langle 1, m\sqrt{d} \right\rangle$. Si $d \not\equiv 1 \pmod 4$, este módulo es $\mathcal{O}_m$ y, en caso contrario,

$$\left\langle 1, m\sqrt{d} \right\rangle = \left\langle 1, m + m\sqrt{d} \right\rangle = \langle 1, 2m\omega \rangle = \mathcal{O}_{2m}. \qquad \blacksquare$$

Ejemplo En la página 404 hemos visto que hay dos clases de equivalencia estrictas de formas cuadráticas de discriminante 56, representadas por las formas $x^2 - 14y^2$, $14x^2 - y^2$. Dichas formas no son equivalentes, pues una toma valores que son restos cuadráticos módulo 7 y la otra sólo toma valores que son restos no cuadráticos módulo 7, por lo que también hay dos clases de equivalencia no estricta.

Sin embargo, si $\mathcal{O} = \langle 1, \sqrt{14} \rangle$, es claro que las clases de similitud estricta de módulos son $[\mathcal{O}]$ y $-[\mathcal{O}]$, por lo que ambas clases forman una única clase de similitud no estricta. Así pues, todo módulo con anillo de coeficientes $\mathbb{Z}[\sqrt{14}]$ es similar a $\mathcal{O}$. Veremos que esto es equivalente a que $\mathbb{Z}[\sqrt{14}]$ es un dominio de factorización única. $\blacksquare$

Similitud de módulos de órdenes reales Hemos visto que para determinar si dos módulos son estrictamente similares basta ver si sus formas cuadráticas asociadas son estrictamente equivalentes, pero para el caso de los módulos asociados a órdenes de cuerpos cuadráticos reales hay un criterio mucho más simple para determinar si son similares o no:

Teorema 12.22 *Sean $M = \langle 1, \gamma\rangle$ y $M' = \langle 1, \gamma'\rangle$ dos módulos con un mismo anillo de coeficientes de discriminante positivo (lo que implica en particular que son módulos en un cuerpo cuadrático real, luego $\gamma, \gamma' \in \mathbb{R}$). Entonces M y M' son similares si y sólo si los desarrollos de γ y γ' en fracción continua son finalmente iguales.*

Demostración: Según 11.15, los desarrollos de γ y γ' son finalmente iguales si y sólo si existen enteros p, q, r, s tales que[2] $ps - qr = \pm 1$ y

$$\gamma' = \frac{p\gamma + q}{r\gamma + s}.$$

Si se cumple esto,

$$\langle 1, \gamma'\rangle = \frac{1}{r\gamma + s}\,\langle r\gamma + s,\ p\gamma + q\rangle = \frac{1}{r\gamma + s}\,\langle 1, \gamma\rangle\,,$$

luego M y M' son similares. Recíprocamente, si existe un ξ tal que $\langle 1, \gamma\rangle = \xi\,\langle 1, \gamma'\rangle$, entonces $\xi\gamma' = p\gamma + q$, $\xi = r\gamma + s$, donde p, q, r, s son enteros racionales tales que $ps - qr = \pm 1$. Ahora basta dividir ambas ecuaciones. ∎

Ejemplo Vamos a determinar si los módulos

$$M = \left\langle 191, 34 + \sqrt{10}\right\rangle, \qquad M' = \left\langle 1, \sqrt{10}\right\rangle$$

son similares. Para ello basta ver si lo son

$$\left\langle 1, \frac{34 + \sqrt{10}}{191}\right\rangle, \qquad \left\langle 1, \sqrt{10}\right\rangle,$$

y basta aplicar el teorema anterior. Consideramos los desarrollos en fracción continua:

$$\frac{34 + \sqrt{10}}{191} = [0, 5, 7, \bar{6}] \qquad \sqrt{10} = [3, \bar{6}],$$

que prueban que ambos módulos son similares. Explícitamente, si llamamos $\alpha = [\bar{6}]$, tenemos que

n	-1	0	1	2	3
a_n	$-$	0	5	7	α
p_n	1	0	1	7	$7\alpha + 1$
q_n	0	1	5	36	$36\alpha + 5$

n	-1	0	1
a_n	$-$	3	α
p_n	1	3	$3\alpha + 1$
q_n	0	1	α

[2] La prueba de 11.15 muestra cómo determinar el signo de $ps - qr$ a partir de los desarrollos en fracción continua, luego en realidad podemos saber si los módulos son o no estrictamente similares.

de donde

$$\frac{34+\sqrt{10}}{191} = \frac{7\alpha+1}{36\alpha+5}, \qquad \alpha = \frac{1}{\sqrt{10}-3} = 3+\sqrt{10}.$$

Multiplicando las matrices

$$\begin{pmatrix} 7 & 1 \\ 36 & 5 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 5 & 21 \end{pmatrix}$$

De aquí se sigue que

$$\frac{34+\sqrt{10}}{191} = \frac{\sqrt{10}+4}{5\sqrt{10}+21},$$

luego, según el teorema anterior,

$$\left\langle 1, \frac{34+\sqrt{10}}{191}\right\rangle = \frac{1}{5\sqrt{10}+21}\left\langle 1, \sqrt{10}\right\rangle,$$

y así

$$\left\langle 191, 34+\sqrt{10}\right\rangle = (21-5\sqrt{10})\left\langle 1, \sqrt{10}\right\rangle,$$

con lo que concluimos que los módulos dados son estrictamente similares. ∎

12.3 Producto de módulos

Introducimos ahora el último ingrediente que necesitamos para estar en condiciones de resolver las ecuaciones diofánticas asociadas a formas cuadráticas irreducibles, y es el hecho de que podemos definir un producto de módulos completos:

Definición 12.23 Si M y M' son dos módulos de un cuerpo cuadrático k, definimos su producto como el conjunto MM' formado por todos los elementos de k de la forma

$$m_1m_1' + \cdots + m_rm_r',$$

donde r es un número natural, los elementos m_i están en M y los m_i' en M'.

El producto de módulos es un módulo, pues si $M = \langle\alpha, \beta\rangle$ y $M' = \langle\gamma, \delta\rangle$, entonces $MM' = \langle\alpha\gamma, \alpha\delta, \beta\gamma + \beta\delta\rangle$.

En efecto, cada $m_i = a_i\alpha + b_i\beta$ y cada $m_i' = c_i\gamma + d_i\delta$, luego

$$m_im_i' = a_ic_i\alpha\gamma + a_id_i\alpha\delta + b_ic_i\beta\gamma + b_id_i\beta\delta,$$

lo que prueba que $MM' \subset \langle\alpha\gamma, \alpha\delta, \beta\gamma, \beta\delta\rangle$. La inclusión opuesta es inmediata, pues todo elemento del módulo de la derecha es de la forma

$$(a\alpha)\gamma + (b\alpha)\delta + (c\beta)\gamma + (d\beta)\delta,$$

que claramente es de la forma requerida por la definición de MM'.

Vamos a ver ahora que el producto de módulos completos es un módulo completo. Con la notación anterior,

$$M = \alpha\,\langle 1, \beta'\rangle\,, \quad M' = \gamma\,\langle 1, \delta'\rangle\,,$$

donde $\beta' = \beta/\alpha$, $\delta' = \delta/\gamma$ son números irracionales. Es fácil ver entonces que

$$MM' = \alpha\gamma(\langle 1, \beta'\rangle\,\langle 1, \delta'\rangle),$$

luego basta ver que el último producto de módulos, es decir,

$$N = \langle 1, \beta', \delta', \beta'\delta'\rangle$$

es un módulo completo. Si no lo fuera, $N = \langle\xi\rangle$, para cierto ξ, de modo que los elementos de N serían de la forma $m\xi$, con m entero. Si ξ fuera irracional, todos los elementos no nulos de N serían irracionales, en contra de que $1 \in N$, mientras que si ξ fuera racional, todos los elementos de N serían racionales, en contra de que $\beta' \in N$.

Ejemplo Vamos a calcular el producto de los módulos $M = \langle 3, 5 + 2\sqrt{7}\rangle$ y $M' = \langle 6, -1 + 3\sqrt{7}\rangle$. En principio es

$$MM' = \left\langle 18, -3 + 9\sqrt{7}, 30 + 12\sqrt{7}, 37 + 13\sqrt{7}\right\rangle,$$

y ahora podemos eliminar generadores empleando la técnica descrita en la prueba del teorema 12.2:

$$\begin{pmatrix} -3 & 9 \\ 18 & 0 \\ 30 & 12 \\ 37 & 13 \end{pmatrix} \sim \begin{pmatrix} -3 & 9 \\ 18 & 0 \\ 33 & 3 \\ 40 & 4 \end{pmatrix} \sim \begin{pmatrix} 33 & 3 \\ -102 & 0 \\ 18 & 0 \\ 7 & 1 \end{pmatrix} \sim \begin{pmatrix} 7 & 1 \\ 12 & 0 \\ 6 & 0 \\ 6 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 \\ 6 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Concluimos que $MM' = \langle 6, 1 + \sqrt{7}\rangle$. ∎

El producto de módulos completos es asociativo, pues si tenemos tres módulos M, M' y M'', es fácil ver que $MM'M''$ está formado por los elementos de la forma

$$m_1 n_1' m_1'' + \cdots + m_r m_r' m_r'',$$

donde r es un número natural, los elementos m_i están en M, los m_i' en M' y los m_i'' en M''. También es obvio que es conmutativo. Ahora probamos un resultado fundamental:

Teorema 12.24 *Si M es un módulo completo con anillo de coeficientes $\mathcal{O}$ y $\bar{M}$ es su módulo conjugado, entonces $M\bar{M} = \mathrm{N}(M)\mathcal{O}$.*

Demostración: Supongamos en primer lugar que el módulo M es de la forma $M = \langle 1, \gamma\rangle$. Entonces, con la notación del teorema 12.9, tenemos que

$$M\bar{M} = \langle 1, \gamma, \bar{\gamma}, \gamma\bar{\gamma}\rangle = \left\langle 1, \gamma, -\gamma - \frac{b}{a}, -\frac{c}{a}\right\rangle = \frac{1}{a}\,\langle a, b, c, a\gamma\rangle\,,$$

donde hemos usado que

$$(x-\gamma)(x-\bar{\gamma}) = x^2 + \frac{b}{a}x + \frac{c}{a},$$

luego

$$-\frac{b}{a} = \gamma + \bar{\gamma}, \qquad \frac{c}{a} = \gamma\bar{\gamma}.$$

Puesto que $(a,b,c) = 1$, aplicando dos veces la relación de Bezout obtenemos que $1 = ua + vb + wc$, para ciertos enteros u, v, w, por lo que

$$M\bar{M} = \frac{1}{a}\langle 1, a\gamma\rangle = \mathrm{N}(M)\mathcal{O},$$

pues tras la definición 12.12 hemos visto que $\mathrm{N}(M) = 1/a$.

Si M es un módulo arbitrario, entonces $M = \alpha M'$, donde M' tiene la forma anterior, luego

$$M\bar{M} = \alpha\bar{\alpha}M'\bar{M}' = \mathrm{N}(\alpha)\,\mathrm{N}(M')\mathcal{O} = |\,\mathrm{N}(\alpha)|\,\mathrm{N}(M')\mathcal{O} = \mathrm{N}(M)\mathcal{O}. \qquad \blacksquare$$

De aquí obtenemos muchas consecuencias:

Teorema 12.25 *Si $\mathcal{O}$ es un orden de un cuerpo cuadrático k, el conjunto de todos los módulos completos de k cuyo anillo de coeficientes es $\mathcal{O}$ es un grupo abeliano con el producto que hemos definido.*

Demostración: Sean M y M' dos módulos con anillo de coeficientes $\mathcal{O}$ y llamemos $\mathcal{O}'$ al anillo de coeficientes de MM'. Entonces

$$\mathrm{N}(MM')\mathcal{O}' = MM'\bar{M}\bar{M}' = \mathrm{N}(M)\mathcal{O}\,\mathrm{N}(M')\mathcal{O} = \mathrm{N}(M)\,\mathrm{N}(M')\mathcal{O},$$

donde hemos usado que $\mathcal{O}\mathcal{O} = \mathcal{O}$, pues la inclusión $\mathcal{O}\mathcal{O} \subset \mathcal{O}$ se debe a que $\mathcal{O}$ es un anillo, luego al sumar y multiplicar elementos de $\mathcal{O}$ obtenemos de nuevo un elemento de $\mathcal{O}$, mientras que la inclusión $\mathcal{O} \subset \mathcal{O}\mathcal{O}$ se debe a que todo elemento de $\mathcal{O}$ es de la forma $\alpha = \alpha\cdot 1 \in \mathcal{O}\mathcal{O}$.

Esto significa que $\mathcal{O}$ y $\mathcal{O}'$ son similares, luego ambos tienen el mismo anillo de coeficientes, pero cada uno es su propio anillo de coeficientes, así que $\mathcal{O} = \mathcal{O}'$.

Por lo tanto, el producto de módulos con anillo de coeficientes $\mathcal{O}$ tiene anillo de coeficientes $\mathcal{O}$.

Por otra parte, $\mathcal{O}$ es el elemento neutro para el producto, pues $M\mathcal{O} = M$. En efecto, la inclusión $M\mathcal{O} \subset M$ se debe a que los elementos de $\mathcal{O}$ son coeficientes de M, luego al multiplicar elementos de M por elementos de $\mathcal{O}$ y sumar los resultados obtenemos elementos de M. La inclusión opuesta se debe a que todo elemento de M es de la forma $\alpha = \alpha\cdot 1 \in M\mathcal{O}$.

Finalmente, el teorema anterior implica que el módulo

$$M^{-1} = \mathrm{N}(M)^{-1}\bar{M}$$

es el inverso de M. (Notemos que, al ser similar a M, también tiene a $\mathcal{O}$ por anillo de coeficientes.) ∎

En la prueba del teorema anterior hemos visto que si M y M' tienen anillo de coeficientes $\mathcal{O}$, entonces

$$\mathrm{N}(MM')\mathcal{O} = \mathrm{N}(M)\,\mathrm{N}(M')\mathcal{O},$$

de donde, tomando normas (por las propiedades 1 y 2 demostradas tras la definición 12.12), resulta que $\mathrm{N}(MM') = \mathrm{N}(M)\,\mathrm{N}(M')$. En suma, hemos probado que la norma de módulos es multiplicativa:

Teorema 12.26 *Si M y M' son dos módulos de un cuerpo cuadrático con el mismo anillo de coeficientes, entonces*

$$\mathrm{N}(MM') = \mathrm{N}(M)\,\mathrm{N}(M').$$

Más aún, si tenemos dos pares de módulos (estrictamente) similares con el mismo anillo de coeficientes $\mathcal{O}$, digamos $M = \alpha M'$, $N = \beta N'$, entonces

$$MN = \alpha\beta M'N',$$

luego MN es (estrictamente) similar a $M'N'$ y por consiguiente podemos definir un producto en el conjunto de las clases de similitud (estricta) de módulos con anillo de coeficientes $\mathcal{O}$ mediante

$$[M][N] = [MN],$$

sin que importen los representantes elegidos en cada clase.

Es claro que con este producto el conjunto de clases de similitud (estricta) de módulo con anillo de coeficientes $\mathcal{O}$ se convierte en un grupo abeliano finito, llamado *grupo de clases (estrictas)* de $\mathcal{O}$.

Observemos que la clase inversa de una clase $[M]$ es simplemente la clase conjugada $[M]^{-1} = [\bar{M}]$, pues

$$[M][\bar{M}] = [M\bar{M}] = [\mathrm{N}(M)\mathcal{O}] = [\mathcal{O}] = 1.$$

Ejemplo Hemos visto antes que hay 6 clases de similitud estricta de módulos correspondientes al orden de discriminante 316, que podemos representar así:

$$\pm 1, \qquad \pm\mathfrak{m}, \qquad \pm\overline{\mathfrak{m}}.$$

Aquí llamamos $1 = [\mathcal{O}] = [\langle 79, \sqrt{79}\rangle$ a la clase principal y $-1 = [\sqrt{79}\mathcal{O}]$, mientras que $\mathfrak{m} = [\langle 3, 1 + \sqrt{79}\rangle]$.

Notemos que, en general, la clase que hemos llamado $-[M] = [\sqrt{79}M]$ no es sino $(-1)[M]$, pues

$$-[M] = [\sqrt{79}M] = [\sqrt{79}\mathcal{O}M] = [\sqrt{79}\mathcal{O}][M] = (-1)[M].$$

Otro hecho obvio es que $(-1)^2 = 1$, pues se trata de $(-1)^2 = [79\mathcal{O}] = [\mathcal{O}]$. Ahora calculamos

$$\mathfrak{m}^2 = [\left\langle 9, 3 + 3\sqrt{79}, 80 + 2\sqrt{79}\right\rangle].$$

Reducimos los generadores:

$$\begin{pmatrix} 80 & 2 \\ 3 & 3 \\ 9 & 0 \end{pmatrix} \sim \begin{pmatrix} 80 & 2 \\ -77 & 1 \\ 9 & 0 \end{pmatrix} \sim \begin{pmatrix} -77 & 1 \\ 234 & 0 \\ 9 & 0 \end{pmatrix} \sim \begin{pmatrix} 4 & 1 \\ 9 & 0 \\ 0 & 0 \end{pmatrix},$$

con lo que

$$\mathfrak{m}^2 = [\left\langle 9, 4 + \sqrt{79}\right\rangle].$$

La forma cuadrática asociada a esta base (que está positivamente orientada) es

$$9x^2 + 8xy - 7y^2,$$

que es primitiva y es estrictamente equivalente a $10x^2 + 14xy - 3y^2$, que es la forma asociada a la clase $-\overline{\mathfrak{m}}$. Así pues, concluimos que $\mathfrak{m}^2 = -\overline{\mathfrak{m}}$, de donde se sigue que $\mathfrak{m}^3 = -\overline{\mathfrak{m}}\mathfrak{m} = -1$, luego $\mathfrak{m}^4 = -\mathfrak{m}$, $\mathfrak{m}^5 = -\mathfrak{m}^2 = \overline{\mathfrak{m}}$ y $\mathfrak{m}^6 = \mathfrak{m}\overline{\mathfrak{m}} = 1$.

En resumen, concluimos que el grupo de clases es cíclico y que un generador es $\mathfrak{m}$. Resumimos los cálculos en una tabla:

n	0	1	2	3	4	5
$\mathfrak{m}^n$	1	$\mathfrak{m}$	$-\overline{\mathfrak{m}}$	-1	$-\mathfrak{m}$	$\overline{\mathfrak{m}}$

Si consideramos clases de similitud no estrictas, entonces $-1 = 1$, con lo que tenemos las clases $1, \mathfrak{m}$ y $\overline{\mathfrak{m}}$, que forman también un grupo cíclico. ∎

Ejercicio: Determinar la estructura del grupo de clases de similitud estricta de módulos correspondientes al orden cuadrático de discriminante 328 y 540.

Terminamos esta sección con una observación elemental:

Teorema 12.27 *El producto de dos módulos enteros de un orden cuadrático es un módulo entero.*

Demostración: Si $\mathfrak{m}$ y $\mathfrak{n}$ son dos módulos enteros con anillo de coeficientes $\mathcal{O}$, sabemos que $\setminus$ tiene anillo de coeficientes $\mathcal{O}$, y sólo hay que probar que $\mathfrak{mn} \subset \mathcal{O}$. Ahora bien, si $\alpha \in \mathfrak{m}$ y $\beta \in \mathfrak{n}$, entonces $\alpha\beta \in \mathfrak{m} \subset \mathcal{O}$, porque los elementos de $\mathfrak{n}$ son coeficientes de $\mathfrak{m}$ y todo elemento de $\setminus$ es suma de productos $\alpha\beta$, luego está en $\mathcal{O}$. ∎

12.4 Ecuaciones diofánticas definidas por formas cuadráticas

Finalmente estamos en condiciones de resolver las ecuaciones diofánticas definidas por formas cuadráticas irreducibles (primitivas y definidas positivas si

su discriminante es negativo). Retomando el ejemplo planteado al principio del capítulo:

$$2x^2 + 22xy - 7y^2 = 77,$$

sabemos que sus soluciones enteras se corresponden con los elementos del módulo $M = \langle 2, 11 + 3\sqrt{15}\rangle$ de norma 154. Con esto no parece que hayamos adelantado mucho, porque encontrar tales elementos es exactamente lo mismo que encontrar las soluciones de la ecuación. Sin embargo, ahora podemos hacer una manipulación que cambia completamente el enfoque del problema.

Vamos a plantear la situación general: tenemos una módulo $M = \langle \alpha, \beta\rangle$ y una ecuación diofántica

$$f(x, y) = \frac{\mathrm{N}(x\alpha + y\beta)}{\mathrm{N}(M)} = m.$$

Cambiando f por $-f$ si es preciso, podemos suponer que $m > 0$ (esto si el discriminante de f es positivo, pero si es negativo y f es definida positiva es necesario que $m > 0$ para que la ecuación tenga solución). Sus soluciones se corresponden con los elementos de M de norma $m\,\mathrm{N}(M)$. Si ξ es uno de estos elementos, entonces el módulo $\mathfrak{m} = \xi M^{-1}$ cumple que

$$\mathfrak{m}M = \xi\mathcal{O} \subset M$$

(pues ξ está en M y los elementos de $\mathcal{O}$ son coeficientes de M), pero esto significa también que todos los elementos de $\mathfrak{m}$ son coeficientes de M, luego $\mathfrak{m} \subset \mathcal{O}$ y así $\mathfrak{m}$ es un módulo entero. Su norma es $\mathrm{N}(\mathfrak{m}) = \mathrm{N}(\xi)\,\mathrm{N}(M)^{-1} = m$ y además, si $C = [M]$ es la clase de equivalencia estricta de M, como $m > 0$, tenemos que $\mathfrak{m}$ es estrictamente similar a M^{-1}, luego $\mathfrak{m} \in C^{-1}$.

Recíprocamente, si $\mathfrak{m} \in C^{-1}$ es un módulo entero de norma m, existe un ξ de norma positiva tal que $\mathfrak{m} = \xi M^{-1}$, luego $\xi \in \mathfrak{m}M \subset M$, porque $\mathfrak{m}$ es entero, luego está contenido en su anillo de coeficientes $\mathcal{O}$, que es también el de M. Además $\mathrm{N}(\xi) = m\,\mathrm{N}(M)$.

Así pues, vemos que los elementos $\xi \in M$ de norma $m\,\mathrm{N}(M)$ se corresponden con los módulos enteros $\mathfrak{m} \in C^{-1}$ de norma m, por la relación $\mathfrak{m} = \xi M^{-1}$.

Notemos que dos elementos ξ y ξ' se corresponden con el mismo módulo $\mathfrak{m}$ si y sólo si $\mathfrak{m} = \xi M^{-1} = \xi' M^{-1}$, y entonces $M^{-1} = \xi^{-1}\xi' M^{-1}$, luego $\epsilon = \xi^{-1}\xi'$ es un coeficiente de M^{-1}, es decir, $\epsilon \in \mathcal{O}$ y $\mathrm{N}(\epsilon) = 1$, luego ϵ es una unidad de $\mathcal{O}$ de norma 1 tal que $\xi' = \epsilon\xi$. Recíprocamente, es claro que si $\xi' = \epsilon\xi$, entonces ξ y ξ' se corresponden con el mismo módulo entero $\mathfrak{m} = \xi' M^{-1} = \xi\epsilon M^{-1} = \xi M^{-1}$.

El teorema siguiente recoge lo que hemos probado:

Teorema 12.28 *Sea $\mathcal{O}$ un orden cuadrático de discriminante D, sea*

$$f(x, y) = \frac{\mathrm{N}(x\alpha + y\beta)}{\mathrm{N}(M)}$$

una forma cuadrática de discriminante D, donde $M = \langle \alpha, \beta\rangle$. Sea C la clase de similitud estricta de M y sea $m > 0$ un número natural. Entonces:

1. *Entonces las soluciones enteras de la ecuación $f(x,y) = m$ se corresponden biunívocamente con los elementos $\xi = x\alpha + y\beta \in M$ de norma $m\,\mathrm{N}(M)$.*

2. *Estos ξ se corresponden con los módulos enteros $\mathfrak{m} = \xi M^{-1} \in C^{-1}$ de norma m.*

3. *Si $\xi_0 \in M$ se corresponde con un módulo $\mathfrak{m}$, los demás valores de ξ que se corresponden con ese mismo M son los de la forma $\xi = \xi_0\epsilon$, donde ϵ recorre las unidades de $\mathcal{O}$ de norma 1.*

4. *Más explícitamente (cuando $D > 0$), si ϵ es la unidad fundamental de $\mathcal{O}$ (o su cuadrado, si ésta tiene norma negativa), los valores de ξ correspondientes a $\mathfrak{m}$ son los de la forma $\xi_n = \xi_0\epsilon^n$, donde n recorre los números enteros.*

Así, con este teorema, para resolver una ecuación $f(x,y) = m$, pasamos de buscar elementos de una norma determinada en un módulo determinado (que no es más fácil que buscar las soluciones enteras de la ecuación) a buscar módulos enteros de una determinada norma en una determinada clase de similitud estricta.

Vamos a ver que este enfoque más abstracto nos permite encontrar en la práctica las soluciones buscadas, pero antes observemos que el teorema anterior se simplifica si queremos únicamente un criterio de existencia de solución, es decir, un criterio que determine si un número m puede expresarse en términos de una forma cuadrática dada (por ejemplo, si es suma de dos cuadrados, etc.):

Teorema 12.29 *Sea $f(x,y) = m$ una ecuación diofántica determinada por una forma cuadrática, donde $m > 0$, y sea C la clase de similitud estricta de módulos asociada a $[f]$. Entonces la ecuación tiene solución si y sólo si la clase C^{-1} contiene módulos enteros de norma m.*

Sucede que, aunque el concepto de "módulo entero de norma m" es mucho más abstracto que el de "elemento de norma m", es más fácil encontrar los primeros que los segundos.

En efecto, por el teorema 12.16, todo módulo entero de norma m con anillo de coeficientes $\mathcal{O}$ de discriminante D es de la forma $\mathfrak{m} = ka^2\langle 1, \gamma\rangle$, donde k y a son números naturales tales que $m = \mathrm{N}(\mathfrak{m}) = k^2a$, luego $\mathfrak{m}' = \langle 1, \gamma\rangle$ tiene norma $\mathrm{N}(\mathfrak{m}') = 1/a$. Por el teorema 12.9 y la propiedad 3 tras la definición 12.12 tenemos que γ es raíz de un polinomio $ax^2 + bx + c$, donde $(a,b,c) = 1$ (y a es precisamente el dado por la norma de $\mathfrak{m}'$). El hecho de que el anillo de coeficientes de $\mathfrak{m}'$ sea $\mathcal{O}$ se traduce en que $b^2 - 4ac = D$. Por último, vamos a ver que podemos exigir que $-a < b \le a$.

Esto se debe a que $\langle 1, \gamma\rangle = \langle 1, \gamma + s\rangle$, para cualquier entero s, y el polinomio correspondiente a $\gamma + s$ es fácil de calcular, pues

$$a(\gamma+s)^2 = a\gamma^2 + 2as\gamma + as^2 = -b\gamma - c + 2as\gamma + as^2$$

$$= (-b + 2as)\gamma + as^2 - c = (-b + 2as)(\gamma + s) + bs - as^2 - c,$$

luego $\gamma + s$ es raíz de

$$ax^2 + (b - 2as)x - bs + as^2 + c,$$

cuyos coeficientes son primos entre sí. Por lo tanto, eligiendo s, podemos hacer que b cumpla lo requerido.

Con esto tenemos un proceso finito para encontrar todos los módulos enteros de una norma dada m correspondientes al orden cuadrático de discriminante D:

1. Consideramos todas las descomposiciones posibles $m = k^2a$.
2. Para cada una de ellas, buscamos todos los valores de b y c que cumplen que $(a, b, c) = 1$, $-a < b \le a$ y $b^2 - 4ac = D$.
3. Para cada una de ellas consideramos

$$\gamma = \frac{-b + \sqrt{D}}{2a}$$

y formamos el módulo

$$\mathfrak{m} = ka\,\langle 1, \gamma\rangle = k\left\langle a, \frac{-b + \sqrt{D}}{2}\right\rangle.$$

Notemos que basta tomar una raíz cuadrada, por ejemplo la positiva si $D > 0$ o la de parte imaginaria positiva si $D < 0$, pues la otra raíz corresponde al valor $-\gamma$ que se obtiene con $-b$, salvo si $b = a$, en cuyo caso con la raíz opuesta obtenemos el módulo

$$k\left\langle a, \frac{-a - \sqrt{D}}{2}\right\rangle = k\left\langle a, \frac{a + \sqrt{D}}{2}\right\rangle = k\left\langle a, \frac{-a + \sqrt{D}}{2}\right\rangle,$$

donde en el último paso al segundo generador le hemos restado el primero, y así concluimos que ambas raíces dan lugar al mismo módulo.

Hemos probado que así obtenemos necesariamente todos los módulos enteros de norma m, pero podemos probar que no necesitaremos descartar ningún $\mathfrak{m}$, es decir, que todos los módulos construidos así son enteros de norma m.

En efecto, si $D \equiv 1 \pmod 4$ entonces $D = m^2d$, con $d \equiv 1 \pmod 4$ y d es impar. Como $D = b^2 - 4ac$, también b es impar. Entonces

$$k\left\langle a, \frac{-b + \sqrt{D}}{2}\right\rangle = k\left\langle a, \frac{-b + m\sqrt{d}}{2}\right\rangle = k\left\langle a, \frac{-b - m}{2} + m\frac{1 + \sqrt{d}}{2}\right\rangle$$

y basta aplicar el teorema 12.16, pues

$$a \mid ac = \mathrm{N}\left(\frac{-b + \sqrt{D}}{2}\right).$$

Si $D \equiv 0 \pmod 4$, entonces $D = (2m)^2 d$ y b es par. Así:

$$k\left\langle a, \frac{-b+\sqrt{D}}{2}\right\rangle = k\left\langle a, \frac{-b+2m\sqrt{d}}{2}\right\rangle = k\left\langle a, -\frac{b}{2}+m\sqrt{d}\right\rangle$$

y concluimos igualmente.

Por 12.9, el anillo de coeficientes es

$$\left\langle 1, \frac{-b+\sqrt{D}}{2}\right\rangle,$$

y los mismos casos anteriores, con $k = a = 1$, nos dan que se trata del orden $\langle 1, m\omega\rangle = \mathcal{O}_m$.

Ejemplo Vamos a encontrar los módulos enteros de norma $77 = 7 \cdot 11$ correspondientes al orden de discriminante $D = 540 = 2^2 \cdot 3^3 \cdot 5$, que es el orden $\mathcal{O}_3$ de $k = \mathbb{Q}\big(\sqrt{15}\,\big)$.

1. Como 77 es libre de cuadrados, tiene que ser $a = 77$.

2. Los valores posibles de (a, b, c) que cumplen 2 son:

$$(77, -76, 17), \qquad (77, -34, 2), \qquad (77, 34, 2), \qquad (77, 76, 17).$$

3. Esto nos da cuatro módulos enteros:

$$\mathfrak{m} = \left\langle 77, 38+3\sqrt{15}\right\rangle, \quad \left\langle 77, 17+3\sqrt{15}\right\rangle,$$

$$\left\langle 77, -38+3\sqrt{15}\right\rangle, \quad \left\langle 77, -17+3\sqrt{15}\right\rangle.$$

Para resolver la ecuación diofántica que hemos considerado al inicio de este tema y de de nuevo al inicio de esta sección, tenemos que quedarnos únicamente con los módulos que están en la clase inversa de la clase de $M = \langle 2, 11+3\sqrt{15}\rangle$, es decir, los que son estrictamente similares a $\overline{M} = \langle 2, -11+3\sqrt{15}\rangle$. Podemos identificarlos calculando las formas cuadráticas asociadas y reduciéndolas:

$\langle 2, -11+3\sqrt{15}\rangle$	$2x^2-22xy-7y^2$	$2x^2+22x-7y^2$
$\langle 77, 38+3\sqrt{15}\rangle$	$77x^2+76xy+17y^2$	$2x^2+22x-7y^2$
$\langle 77, 17+3\sqrt{15}\rangle$	$77x^2+34xy+2y^2$	$2x^2+22x-7y^2$
$\langle 77, -38+3\sqrt{15}\rangle$	$77x^2-76xy+17y^2$	$2x^2+22x-7y^2$
$\langle 77, -17+3\sqrt{15}\rangle$	$77x^2-34xy+2y^2$	$2x^2+22x-7y^2$

Vemos así que los cuatro módulos están en la clase correcta. De hecho, observemos que $M = \overline{M}$, pues si al segundo generador de $\overline{M}$ le sumamos 11 veces el primero obtenemos la base de M. ∎

Resolución de $2x^2+22xy-7y^2=77$ Ya estamos en condiciones de resolver la ecuación diofántica que habíamos planteado al principio de este capítulo:

$$2x^2 + 22xy - 7y^2 = 77.$$

Acabamos de dar el primer paso, que es encontrar todos los módulos enteros de norma 77 estrictamente similares a $\overline{M} = M$. Ahora tenemos que encontrar números ξ de norma positiva tales que

$$\mathfrak{m} = \xi M^{-1} = (\xi/2)\overline{M} = (\xi/2)M,$$

para cada uno de los cuatro módulos que hemos encontrado. En la prueba del teorema 12.19 hemos visto cómo hacerlo. Basta encontrar bases $\mathfrak{m} = \langle \alpha', \beta' \rangle$ de los cinco módulos de la tabla que correspondan a la misma forma cuadrática, por ejemplo a $2x^2 + 22x - 7y^2$, que en el caso de M es $(\alpha, \beta) = (2, 11 + 3\sqrt{15})$, y entonces sirve $\xi/2 = \alpha'/\alpha = \alpha'/2$.

Una base de $\mathfrak{m} = \langle 77, 38 + 3\sqrt{15} \rangle$ asociada a la forma $2x^2 + 22xy - 7y^2$ la hemos calculado en el primer apartado de la sección 12.2, y ha resultado ser

$$(\alpha', \beta') = (37 + 9\sqrt{15}, 406 + 105\sqrt{15}),$$

luego, según la prueba del teorema 12.19, nos sirve $\xi = 37 + 9\sqrt{15}$. Ahora calculamos las coordenadas de ξ en la base de M:

$$\xi = 37 + 9\sqrt{15} = x2 + y(11 + 3\sqrt{15}).$$

esto equivale a un sistema de ecuaciones cuya solución es $(x, y) = (2, 3)$. Con esto hemos encontrado una primera solución de la ecuación diofántica dada.

Podríamos repetir el proceso con el módulo $\mathfrak{m} = \langle 77, 17 + 3\sqrt{15} \rangle$, pero en realidad hay un camino más rápido. La solución $(x, y) = (2, 3)$ puede obtenerse fácilmente por tanteo, al igual que la solución $(x, y) = (3, 1)$. A partir de una solución como ésta podemos calcular

$$\xi = 3 \cdot 2 + 1(11 + 3\sqrt{15}) = 17 + 3\sqrt{15},$$

que a su vez corresponde a

$$\begin{aligned} \mathfrak{m} &= \frac{17 + 3\sqrt{15}}{2} \left\langle 2, 11 + 3\sqrt{15} \right\rangle = \left\langle 17 + 3\sqrt{15}, 161 + 42\sqrt{15} \right\rangle \\ &= \left\langle 77, 17 + 3\sqrt{15} \right\rangle, \end{aligned}$$

luego ya tenemos una solución asociada al segundo módulo. Igualmente podríamos haber comprobado que $(2, 3)$ corresponde al primer módulo.

Para los dos módulos que quedan basta observar que son los conjugados de los dos primeros luego, como $M = \overline{M}$, es claro que podemos tomar como ξ los conjugados de los valores que hemos obtenido para los dos primeros módulos.

Concretamente, $\xi = 17 - 3\sqrt{15} = 14 \cdot 2 - (11 + 3\sqrt{15})$ corresponde a la solución $(x, y) = (14, -1)$, mientras que $\xi = 37 - 9\sqrt{15} = 35 \cdot 2 - 3(11 + 3\sqrt{15})$ corresponde a $(x, y) = (35, -3)$.

De este modo hemos obtenido cuatro soluciones de la ecuación:

$$(x, y) = (3, 1), \quad (2, 3), \quad (14, -1), \quad (35, -3),$$

pero no son dos soluciones cualesquiera, sino que el teorema 12.28 nos asegura que cualquier otra se obtiene de estas multiplicando el ξ correspondiente por una unidad de $\mathcal{O}_3 \subset \mathbb{Q}(\sqrt{15})$ de norma 1. En realidad, se cumple que todas las unidades de $\mathcal{O}_3$ tienen norma 1, pues ya habíamos señalado que la unidad fundamental es $\epsilon = 244 + 63\sqrt{15}$. Así pues, las soluciones de la ecuación son las de la forma

$$\begin{aligned} 2x + (11 + 3\sqrt{15})y &= \pm(37 \pm 9\sqrt{15})(244 + 63\sqrt{15})^n, \\ 2x + (11 + 3\sqrt{15})y &= \pm(17 \pm 3\sqrt{15})(244 + 63\sqrt{15})^n, \end{aligned}$$

donde n recorre los números enteros.

También podemos expresar las soluciones de forma recurrente, planteando

$$\begin{aligned} 2x_{n+1} + (11 + 3\sqrt{15})y_{n+1} &= (2x_n + (11 + 3\sqrt{15})y_n)(244 + 63\sqrt{15}) \\ &= (487x_n + 5\,519y_n) + (126x_n + 1\,425y_n)\sqrt{15}. \end{aligned}$$

Al resolver el sistema de ecuaciones que resulta de igualar las coordenadas, obtenemos que

$$x_{n+1} = 13x_n + 147y_n, \qquad y_{n+1} = 42x_n + 475y_n.$$

Si en lugar de ϵ usamos ϵ^{-1} obtenemos:

$$x_{n+1} = 475x_n - 147y_n, \qquad y_{n+1} = -42x_n + 13y_n.$$

En total tenemos ocho sucesiones de soluciones, obtenidas por las dos relaciones de recurrencia anteriores partiendo de las cuatro soluciones fundamentales (x_0, y_0) que hemos encontrado (más otras ocho con las soluciones de signos opuestos). Los primeros términos son:

-2	-1	0	1	2
$(623\,661, -55\,145)$	$(1\,278, -113)$	$(3, 1)$	$(186, 601)$	$(90\,765, 293\,287)$
$(248\,390, -21\,963)$	$(509, -45)$	$(2, 3)$	$(467, 1\,509)$	$(227\,894, 736\,389)$
$(3\,316\,922, -293\,287)$	$(6\,797, -601)$	$(14, -1)$	$(35, 113)$	$(17\,066, 55\,145)$
$(8\,328\,173, -736\,389)$	$(17\,066, -1\,509)$	$(35, -3)$	$(14, 45)$	$(6\,797, 21\,963)$

■

Notemos que el procedimiento que hemos seguido para resolver la ecuación cuadrática del ejemplo precedente es totalmente general, en el sentido de que se puede aplicar igualmente a cualquier ecuación definida por cualquier forma cuadrática irreducible.

Ejemplo Vamos a encontrar las soluciones enteras de la ecuación

$$2x^2 + 22xy - 7y^2 = 36\,180.$$

El procedimiento es el mismo que acabamos de emplear, así que resumimos los cálculos. Ahora

$$m = 36\,180 = 2^2 \cdot 3^3 \cdot 5 \cdot 67,$$

luego las descomposiciones $m = k^2 a$ pueden tener $k = 1, 2, 3, 6$. La tabla siguiente contiene los únicos valores de (k, a, b, c) que cumplen $-a < b \leq a$, $b^2 - 4ac = 540$ junto con el módulo correspondiente:

k	a	b	c	$\mathfrak{m}$
2	9 045	−6 300	1 097	$2\langle 9\,045, 3\,150 + 3\sqrt{15}\rangle$
2	9 045	−5 760	917	$2\langle 9\,045, 2\,880 + 3\sqrt{15}\rangle$
2	9 045	−270	2	$2\langle 9\,045, 135 + 3\sqrt{15}\rangle$
2	9 045	270	2	$2\langle 9\,045, -135 + 3\sqrt{15}\rangle$
2	9 045	5 760	917	$2\langle 9\,045, -2\,880 + 3\sqrt{15}\rangle$
2	9 045	6 300	1 097	$2\langle 9\,045, -3\,150 + 3\sqrt{15}\rangle$
6	1 005	−270	18	$6\langle 1\,005, 135 + 3\sqrt{15}\rangle$
6	1 005	270	18	$6\langle 1\,005, -135 + 3\sqrt{15}\rangle$

Una forma cuadrática asociada es la que tiene coeficientes $(a, -b, c)$. Las seis primeras son estrictamente equivalentes a $2x^2 + 22xy - 7y^2$, mientras que las dos últimas se reducen a $18x^2 + 18xy - 3y^2$, que corresponde a otra clase de equivalencia, así que podemos descartarlas. La tabla siguiente muestra el cambio de variables que transforma cada una de las seis primeras formas en la forma reducida $2x^2 + 22xy - 7y^2$ junto con la base de $\mathfrak{m}$ asociada a dicha forma:

$x = 115x' + 1\,301y'$ $y = 329x' + 3\,722y'$	$\langle 7\,650 + 1\,974\sqrt{15}, 86\,490 + 22\,332\sqrt{15}\rangle$
$x = -125x' - 1\,414y'$ $y = -391x' - 4\,423y'$	$\langle -9\,090 - 2\,346\sqrt{15}, -102\,780 - 26\,538\sqrt{15}\rangle$
$x = \quad - \quad y'$ $y = x' - 62y'$	$\langle -270 + 6\sqrt{15}, -1\,350 - 372\sqrt{15}\rangle$
$x = \quad - \quad y'$ $y = x' + 73y'$	$\langle 270 + 6\sqrt{15}, 1\,620 + 438\sqrt{15}\rangle$
$x = -13x' - 150y'$ $y = \quad 41x' + 473y'$	$\langle 990 + 246\sqrt{15}, 10\,980 + 2\,838\sqrt{15}\rangle$
$x = -17x' - 195y'$ $y = \quad 49x' + 562y'$	$\langle 1\,170 + 294\sqrt{15}, 13\,050 + 3\,372\sqrt{15}\rangle$

El número ξ que cumple $\mathfrak{m} = \xi M^{-1}$ es el primer elemento de la base de cada módulo, y sus coordenadas en la base de $M = \langle 2, 11 + 3\sqrt{15}\rangle$ son las soluciones

$$(x, y) = (206, 658), \quad (-244, -782), \quad (-146, 2), \quad (124, 2), \quad (44, 82), \quad (46, 98).$$

A partir de cada una de estas soluciones y de sus opuestas podemos generar dos sucesiones de soluciones mediante las recurrencias

$$x_{n+1} = 13x_n + 147y_n, \qquad y_{n+1} = 42x_n + 475y_n.$$

$$x_{n+1} = 475x_n - 147y_n, \qquad y_{n+1} = -42x_n + 13y_n.$$ ∎

Las técnicas que hemos desarrollado no sólo nos permiten resolver ecuaciones cuadráticas, sino que nos permiten demostrar muchas de las conjeturas que habíamos encontrado en el capítulo anterior. Por ejemplo:

Teorema 12.30 *Sean $[f_1]$ y $[f_2]$ dos clases de equivalencia estrictas de formas cuadráticas de discriminante D correspondientes a las clases de similitud estricta de módulos C_1 y C_2, y sea $[f_3]$ la clase correspondiente a C_1C_2. Entonces, el producto de un número de la forma $f_1(x,y)$ por otro de la forma $f_2(x,y)$ es de la forma $f_3(x,y)$.*

DEMOSTRACIÓN: Si las ecuaciones $f_1(x,y) = m_1$ y $f_2(x,y) = m_2$ tienen soluciones enteras, el teorema 12.29 nos da que las clases C_1^{-1} y C_2^{-1} contienen, respectivamente, módulos enteros $\mathfrak{m}_1$ y $\mathfrak{m}_2$ de normas m_1 y m_2. Pero entonces $\mathfrak{m}_1\mathfrak{m}_2 \in C_1^{-1}C_2^{-1} = (C_1C_2)^{-1}$ es un módulo entero de norma m_1m_2, luego, si $[f_3]$ es la clase de formas correspondiente a C_1C_2, tenemos que la ecuación $f_3(x,y) = m_1m_2$ tiene soluciones enteras. ∎

Esto es lo que tal vez el lector haya conjeturado en el capítulo anterior al considerar las formas

$$x^2 + 5y^2, \qquad 2x^2 + 2xy + 3y^2.$$

Son representantes de las clases de equivalencia estricta de formas de discriminante -20, que se corresponden con las dos clases, digamos 1 y C, de módulos con anillo de coeficientes $\mathcal{O} = \mathbb{Z}\big[\sqrt{-5}\,\big]$. Concretamente, la forma principal $x^2 + 5y^2$ se corresponde con el elemento neutro 1 del grupo de clases, y la segunda forma se corresponde con la clase C. Como $1^1 = 1 = C^2$ y $1 \cdot C = C$, hemos demostrado que el producto de dos enteros de la forma $2x^2 + 2xy + 3y^2$ es de la forma $x^2 + 5y^2$, o que el producto de un entero de la forma $x^2 + 5y^2$ por otro de la forma $2x^2 + 2xy + 3y^2$ es de la forma $2x^2 + 2xy + 3y^2$, etc.

Del mismo modo, el grupo de orden 4 que obtuvimos empíricamente en el capítulo anterior a partir de las formas cuadráticas de discriminante $D = 328$ (página 404) no es sino el reflejo del grupo de clases de similitud estricta de los módulos cuyo anillo de coeficientes es el orden cuadrático de dicho discriminante que se da a consecuencia del teorema anterior.

Otro hecho que habíamos constatado sin que supiéramos justificarlo es que un número es de la forma $x^2 + 5y^2$ si y sólo si lo es su parte libre de cuadrados, e igualmente con los números de la forma $2x^2 + 2xy + 3y^2$. Los resultados de este tipo son consecuencia del teorema siguiente:

Teorema 12.31 *Sea k un cuerpo cuadrático, sea $\mathcal{O}_m \subset k$ uno de sus órdenes, sea $\mathfrak{m} = \langle a, b + m\omega\rangle$ un módulo entero con anillo de coeficientes $\mathcal{O}_m$ y $\mathrm{N}(\mathfrak{m}) = a$. Sea $a = ua'$, donde $(a', m) = 1$. Entonces $\mathfrak{m}' = \langle a', b + m\omega\rangle$ también es un módulo entero con anillo de coeficientes $\mathcal{O}_m$ y $\mathrm{N}(\mathfrak{m}') = a'$.*

Demostración: El teorema 12.16 nos da que $a \mid \mathrm{N}(b+m\omega)$, luego también $a' \mid \mathrm{N}(b + m\omega)$, y de nuevo el teorema 12.16 implica que todos los elementos de $\mathcal{O}_m$ son coeficientes de $\mathfrak{m}' \subset \mathcal{O}_m$, luego $\mathfrak{m}'$ es entero. Lo único que podría fallar es que el anillo de coeficientes de $\mathfrak{m}'$ no fuera exactamente $\mathcal{O}_m$, sino un orden mayor $\mathcal{O}_{m'}$. Notemos que para que pueda darse la inclusión $\mathcal{O}_m \subset \mathcal{O}_{m'}$ es necesario que $m' \mid m$, digamos $m = cm'$. Entonces

$$\mathfrak{m}' = \langle a', b + cm'\omega\rangle\,.$$

Sea k el mayor número natural que divide a todos los elementos de $\mathfrak{m}'$ en $\mathcal{O}_{m'}$. Necesariamente $a' = a''k$, $b = b''k$, $c = c''k$ y

$$\mathfrak{m}' = k\,\langle a'', b'' + c''m'\omega\rangle\,.$$

En estas condiciones, en la prueba del teorema 12.16 hemos visto que, para que los elementos de $\mathcal{O}_{m'}$ sean coeficientes de $\mathfrak{m}'$, es necesario que $c'' = 1$, pero entonces $c = k$ y $k \mid (a', m) = 1$, luego $k = 1$, luego $c = 1$, luego $m' = m$. ∎

Notemos que la condición $(a', m) = 1$ se cumple trivialmente si $m = 1$.

No podemos afirmar que si hay un módulo entero de norma a también hay módulos enteros (con el mismo anillo de coeficientes $\mathcal{O}_m$) de norma cualquier divisor de a (primo con m) porque no todo módulo entero es de la forma considerada en el teorema anterior. La expresión general es

$$\mathfrak{m} = k\,\langle a, b + m\omega\rangle\,,$$

que tiene norma $\mathrm{N}(\mathfrak{m}) = k^2a$. Lo que podemos afirmar es que si $c = u^2v$, donde v es libre de cuadrados y $(v, m) = 1$, entonces hay un módulo entero con anillo de coeficientes $\mathcal{O}_m$ de norma c si y sólo si hay uno de norma v, pues dicho módulo será de la forma indicada más arriba con $a = w^2v$ y $k = uw$ y basta aplicar el teorema anterior (la otra implicación es trivial, pues si hay un módulo entero de norma v, basta multiplicarlo por u para tener otro de norma u^2v).

En virtud del teorema 12.29, de aquí se siguen estos hechos:

- *Si la parte libre de cuadrados de un entero a es prima con m, entonces a está representado por una forma cuadrática asociada al orden $\mathcal{O}_m$ de un cuerpo cuadrático si y sólo si lo está su parte libre de cuadrados (aunque no sea necesariamente la misma forma).*
- *Si un entero libre de cuadrados es primo en m, entonces está representado por una forma cuadrática asociada al orden $\mathcal{O}_m$ de un cuerpo cuadrático si y sólo si lo están sus factores primos (aunque la forma puede ser distinta para cada primo).*

La situación es especialmente simple cuando $m = 1$ y más aún si sólo hay una clase de equivalencia estricta de formas. La condición sobre m es necesaria, como muestra el ejemplo siguiente:

Ejemplo Hay dos formas cuadráticas reducidas de discriminante $D = -72$, a saber:

$$x^2 + 18y^2, \qquad 2x^2 + 9y^2.$$

El orden cuadrático de este discriminante es $\mathcal{O}_3 \subset \mathbb{Q}(\sqrt{-2})$. El módulo

$$\mathfrak{m} = \langle 27, 3 + 3\sqrt{-2} \rangle$$

es entero y tiene a $\mathcal{O}_3$ como anillo de coeficientes. Notemos que si hacemos

$$\mathfrak{m} = 3\langle 9, 1 + \sqrt{-2} \rangle,$$

el módulo de la derecha no es entero, pues $9 \nmid \mathrm{N}(1 + \sqrt{-2}) = 3$, por lo que no es cierto que el anillo de coeficientes sea $\mathcal{O}_3$. El hecho de que $\mathrm{N}(\mathfrak{m}) = 27$ se traduce en la representación

$$27 = 3^2 + 18 \cdot 1^2.$$

Sin embargo, es inmediato comprobar que ninguna de las dos formas de discriminante -72 representa a la parte libre de cuadrados de 27, es decir, a 3. Esto equivale a que no existen módulos enteros de norma 3 cuyo anillo de coeficientes sea $\mathcal{O}_3$. El que daría el teorema 12.31, a saber, $\mathfrak{m}' = \langle 3, 3 + 3\sqrt{-2} \rangle$ es ciertamente un módulo entero, pero su anillo de coeficientes es $\mathcal{O}_1$, pues ahora sí que podemos expresarlo como

$$\mathfrak{m}' = 3\langle 1, 1 + \sqrt{-2} \rangle$$

y el módulo de la derecha es entero. ∎

El hecho de que, incluso si $m = 1$, la parte libre de cuadrados puede estar representada por una forma distinta lo ilustra el ejemplo de las formas de discriminante $D = -56$ que consideramos en el capítulo anterior. Vamos a analizarlo a la luz de los resultados que conocemos ahora:

Ejemplo La forma $f(x, y) = x^2 + 14y^2$ es la forma principal de discriminante $D = -56$. El hecho de que, por ejemplo, $f(17, 5) = 3^2 \cdot 71$ se corresponde con que en la clase principal (que es su propia inversa) hay un módulo entero de norma $3^2 \cdot 71 = 639$. Es fácil encontrarlo: Puesto que

$$x^2 + 14y^2 = \mathrm{N}(x + y\sqrt{-14}),$$

la solución $(17, 5)$ se corresponde con el elemento $\xi = 17 + 5\sqrt{-14}$ del módulo $\langle 1, \sqrt{-14} \rangle = \mathcal{O}_1$, luego el módulo entero es

$$\mathfrak{m} = \xi\mathcal{O}_1 = \langle 17 + 5\sqrt{-14}, -70 + 17\sqrt{-14} \rangle = \langle 639, 259 + \sqrt{-14} \rangle.$$

A partir de él podemos obtener, según el teorema 12.31, el módulo entero

$$\mathfrak{m}' = \langle 71, 259 + \sqrt{-14} \rangle = \langle 71, 46 + \sqrt{-14} \rangle,$$

de norma 71, lo que nos asegura que 71 está representado por las formas cuadráticas correspondientes a la clase inversa de la clase de $\mathfrak{m}'$, es decir, la clase de

$$\overline{\mathfrak{m}}' = \left\langle 71, -46 + \sqrt{-14} \right\rangle.$$

La forma asociada a esta base es $71x^2 - 92xy + 30y^2$, que se reduce a $2x^2 + 7y^2$. En efecto, vemos que $71 = 2 \cdot 2^2 + 7 \cdot 3^2$ está representado por esta forma, pero no por la forma principal.

Observemos también que 3 está representado por la forma $3x^2 + 2xy + 5y^2$, lo que se traduce en que existen módulos enteros de norma 3, por ejemplo, $\mathfrak{n} = \left\langle 3, 1 + \sqrt{-14} \right\rangle$, y se comprueba que $\mathfrak{n}^2 = \left\langle 9, 7 + \sqrt{-14} \right\rangle$ y que $\mathfrak{n}^2\mathfrak{m}' = \mathfrak{m}$.

El grupo de clases de similitud estricta del orden $\mathcal{O}_1$ es cíclico de orden 4, y la clase $[\mathfrak{n}]$ es un generador, mientras que $[\mathfrak{n}]^2 = [\mathfrak{m}']$ es la clase de orden 2 y $[\mathfrak{n}^2\mathfrak{m}'] = [\mathfrak{m}]$ es la clase principal, pero no lo es la clase que resulta de eliminar el factor $\mathfrak{n}^2$. ∎

Sin embargo, todavía no estamos en condiciones de entender por qué en el caso de las formas de discriminante -20, es decir,

$$x^2 + 5y^2, \qquad 2x^2 + 2xy + 3y^2,$$

sí que es cierto que un número está representado por una de ellas si y sólo si su parte libre de cuadrados está representada por la misma forma. Esto será inmediato en el capítulo siguiente.

12.5 Ecuaciones diofánticas cuadráticas

Terminamos este capítulo extendiendo los resultados de la sección anterior a la resolución de ecuaciones diofánticas de la forma

$$ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

Llamaremos $D = b^2 - 4ac$ y vamos a considerar en primer lugar el caso en que D no es nulo.

12.5.1 La transformación de Legendre

Vamos a ver que, suponiendo que $D \neq 0$, es posible aplicar un cambio de variables de la forma

$$x' = px + r, \qquad y' = qy + s,$$

donde p, q, r, s son enteros y p, q son no nulos, que transforma cualquier polinomio

$$ax^2 + bxy + cy^2 + dx + ey + f$$

en otro de la forma $ax'^2 + bx'y' + cy'^2 + f'$. Para ello multiplicamos el polinomio por p^2q^2:

$$aq^2(px)^2 + bpq(px)(qy) + cp^2(qy)^2 + dpq^2(px) + ep^2q(qy) + fp^2q^2.$$

Ahora sustituimos $px = x' - r$, $qy = y' - s$ y operamos el resultado es:

$$aq^2x'^2 + bpqx'y' + cp^2y'^2 + q(dpq - 2aqr - bps)x' + p(epq - bqr - 2cps)y'$$
$$+fp^2q^2 + aq^2r^2 + bpqrs + cp^2s^2 - dpq^2r - ep^2qs.$$

Queremos elegir p, q, r, s de modo que

$$2aqr + bps = dpq, \qquad bqr + 2cpqs = epq.$$

Despejando r y s resulta:[3]

$$r = \frac{p(be - 2cd)}{D}, \qquad s = \frac{q(bd - 2ae)}{D}.$$

Por lo tanto, basta tomar $p = q = D$, $r = be - 2cd$, $s = bd - 2ae$.

Concluimos que el cambio de variables

$$x' = Dx + (be - 2cd), \qquad y' = Dy + (bd - 2ae)$$

transforma la ecuación dada en

$$ax'^2 + bx'y' + cy'^2 + fD^2 + ar^2 + brs + cs^2 - dDr - eDs = 0,$$

donde, sustituyendo r y s y operando, se llega a

$$ax'^2 + bx'y' + cy'^2 + fD^2 + fD^2 + D(ae^2 - bde + cd^2) = 0.$$

Con esto hemos demostrado lo siguiente:

Teorema 12.32 *Las soluciones enteras de una ecuación de la forma*

$$ax^2 + bxy + cy^2 + dx + ey + f = 0,$$

donde $D = b^2 - 4ac \neq 0$*, son de la forma*

$$x = \frac{x' + 2cd - be}{D}, \qquad y = \frac{y' + 2ae - bd}{D},$$

donde (x', y') *es una solución entera de la ecuación*

$$ax'^2 + bx'y' + cy'^2 + fD^2 + D(ae^2 - bde + cd^2) = 0$$

que cumpla

$$x' \equiv be - 2cd \pmod{D}, \qquad y' \equiv bd - 2ae \pmod{D}.$$

[3]Aquí es útil aplicar la *regla de Cramer,* según la cual, la solución de un sistema de ecuaciones

$$ax + by = u$$
$$cx + dy = v$$

viene dada por

$$x = \frac{\begin{vmatrix} u & b \\ v & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}, \qquad y = \frac{\begin{vmatrix} a & u \\ c & v \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}$$

Ejemplo Vamos a calcular las soluciones enteras de la ecuación

$$2x^2 + 22xy - 7y^2 - 25x - 2y + 10 = 0,$$

cuyo discriminante es $D = 540$. El teorema anterior, a través del cambio de variables

$$x = \frac{x' + 394}{540}, \qquad y = \frac{y' + 542}{540},$$

reduce el problema a encontrar las soluciones enteras de

$$2x'^2 + 22x'y' - 7y'^2 = 36\,180$$

que cumplen $x' \equiv 146 \pmod{540}$, $y' \equiv 538 \pmod{540}$. Las soluciones enteras de esta ecuación las hemos calculado en la sección anterior, y son las que resultan de aplicar la las seis soluciones siguientes

$$(x', y') = (206, 658), \quad (-244, -782), \quad (-146, 2), \quad (124, 2), \quad (44, 82), \quad (46, 98)$$

y a sus opuestas las relaciones recurrentes

$$x'_{n+1} = 13x'_n + 147y'_n, \qquad y'_{n+1} = 42x'_n + 475y'_n.$$

$$x'_{n+1} = 475x'_n - 147y'_n, \qquad y'_{n+1} = -42x'_n + 13y'_n.$$

De entre ellas debemos seleccionar las que cumplen la congruencia. Para ello observamos que, si llamamos

$$f(x, y) = (13x + 147y, 42x + 475y),$$

si se cumple que $(x, y) \equiv (x', y') \pmod{D}$ (donde esta notación significa que cada coordenada es congruente módulo D con la correspondiente del miembro opuesto) entonces $f(x, y) \equiv f(x', y') \pmod{D}$ (y esto es un hecho general, que no depende que la situación particular de este ejemplo).

Por lo tanto, si partimos de una solución (x'_0, y'_0) y vamos generando la sucesión (x'_n, y'_n) a partir de ella, como los restos módulo D de cada par sólo pueden tomar D^2 valores, al cabo de un número finito de pasos tenemos que encontrar dos índices tales que $(x'_n, y'_n) \equiv (x'_m, y'_m) \pmod{D}$, y a partir de ese momento se repetirá un ciclo de restos módulo D. En realidad, aplicando f^{-1}, de ahí se sigue también que $(x'_{n-1}, y'_{n-1}) \equiv (x'_{m-1}, y'_{m-1}) \pmod{D}$, luego retrocediendo de este modo llegamos a que existe un mínimo índice r tal que $(x'_r, y'_r) \equiv (x'_0, y'_0) \pmod{D}$.

Sólo tenemos que calcular el ciclo correspondiente a cada posible solución inicial y comprobar si en él aparece el valor $(146, 538)$ que nos interesa. Si no es así, ninguna de las soluciones de esa serie proporciona soluciones enteras de la ecuación original.

Por ejemplo, si partimos de $(206, 658)$, el ciclo que obtenemos es

$$(206, 118), (44, 442), (206, 118),$$

en el cual no está el par necesario, luego tenemos que descartar esta sucesión de soluciones. Lo mismo sucede con las 12 soluciones de partida que tenemos (contando las opuestas a las 6 que hemos calculado) excepto con $(146, -2)$, cuyo ciclo es

$$(146, 538), (524, 322), (146, 538).$$

Por lo tanto, vemos que $(146, -2)$ proporciona una solución de la ecuación original, y a partir de ella, podemos obtener otras nuevas aplicando dos veces la transformación f o su inversa. Esto equivale a hacer

$$x'_{n+1} = 6\,343x'_n + 71\,736y'_n, \quad y'_{n+1} = 20\,496x'_n + 231\,799y'_n$$

o

$$x'_{n+1} = 231\,799x'_n - 71\,736y'_n, \quad y'_{n+1=} - 20\,496x'_n + 6\,343y'_n.$$

Finalmente, aplicamos el cambio de variables para obtener la solución

$$(x_0, y_0) = \left(\frac{146 + 394}{540}, \frac{-2 + 542}{540}\right) = (1, 1),$$

y a partir de ella, las demás soluciones se obtienen mediante las relaciones recurrentes

$$(x_{n+1}, y_{n+1}) = (-76\,629 + 6\,343x_n + 71\,736y_n, -247\,611 + 20\,496x_n + 231\,799y_n),$$

$$(x_{n+1}, y_{n+1}) = (-97\,125 + 231\,799x_n - 71\,736y_n, 8\,589 - 20\,496x_n + 6\,343y_n).$$

Por ejemplo, la primera se calcula así:

$$\begin{aligned} x_{n+1} &= \frac{6\,343(540x_n - 394) + 71\,736(540y_n - 542) + 394}{540} \\ &= -76\,629 + 6\,343x_n + 71\,736y_n \\ y_{n+1} &= \frac{20\,496(540x_n - 394) + 231\,799(540y_n - 542) + 542}{540} \\ &= -247\,611 + 20\,496x_n + 231\,799y_n. \end{aligned}$$

Las primeras soluciones son:

n	(x_n, y_n)
-2	$(14\,988\,007\,441, -1\,325\,261\,111)$
-1	$(62\,938, -5\,564)$
0	$(1, 1)$
1	$(1\,450, 4\,684)$
2	$(345\,132\,145, 1\,115\,218\,105)$

■

El ejemplo precedente ilustra cómo la transformación de Legendre considerada en el teorema 12.32 nos permite resolver cualquier ecuación diofántica de dos variables definida por un polinomio de segundo grado con discriminante

$D \neq 0$. Notemos que los casos en que $D < 0$ o D es un cuadrado perfecto no nulo son más simples, porque la forma reducida tiene un número finito de soluciones y, una vez calculadas, basta deshacer el cambio de variables y descartar las que no sean enteras. (Para el caso en que D es cuadrado perfecto, véase el ejemplo de la página 379.)

12.5.2 Ecuaciones con discriminante nulo

Vamos a ver ahora cómo tratar el caso de ecuaciones

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

con discriminante nulo, es decir, tales que $b^2 = 4ac$. Lo ilustraremos con el ejemplo dado por la ecuación

$$27x^2 - 36xy + 12y^2 + 5x - 7y - 5 = 0.$$

Notemos que si $a = c = 0$, entonces la ecuación es lineal y ya hemos visto cómo resolverla.[4] Por simetría podemos suponer $a \neq 0$. Llamamos $g = (a, c)$, $a' = a/g$, $b' = b/g$, $c' = c/g$. Entonces $b'^2 = 4a'c'$, luego b' es entero.

Notemos que, como $ac = b^2 \geq 0$, se cumple que $ac \geq 0$, luego a y c tienen el mismo signo, luego podemos elegir g con el mismo signo que a y c, y así $a', c' \geq 0$. Además, como $(a', c') = 1$ y $(b'/2)^2 = a'c'$, tanto a' como c' son cuadrados perfectos. Digamos que $a' = u^2$ y $c' = v^2$. Podemos tomar $u \geq 0$ y v con el mismo signo que b'. Así $b' = 2uv$.

En nuestro ejemplo $g = (27, -36, 12) = 3$, luego

$$a' = 9 = 3^2, \quad c' = 4 = (-2)^2, \quad b' = -12 = 2 \cdot 3 \cdot (-2).$$

Esto nos permite escribir la ecuación como

$$g(u^2x^2 + 2uvxy + v^2y^2) + dx + ey + f = 0$$

o, equivalentemente,

$$g(ux + vy)^2 + dx + ey + f = 0.$$

Multiplicamos la ecuación por u y sumamos y restamos vdy:

$$ug(ux + vy)^2 + d(ux + vy) + (ue - vd)y + uf = 0.$$

Esto nos permite hacer el cambio de variable $p = ux + vy$, con lo que la ecuación pasa a ser

$$ugp^2 + dp + (ue - vd)y + uf = 0.$$

Equivalentemente,

$$(vd - ue)y = ugp^2 + dp + uf.$$

En nuestro ejemplo es $p = 3x - 2y$, de modo que $11y = 9p^2 + 5p - 15$.

Ahora distinguimos dos casos:

[4] Véase el primer apartado de la sección 2.2.

Si $vd - ue = 0$ Entonces calculamos las raíces enteras p_i del polinomio $ugp^2 + dp + uf$ (que a lo sumo serán dos) y sólo tenemos que resolver las ecuaciones lineales $ux + vy = p_i$.

Si $vd - ue \neq 0$ En este caso dividimos

$$p = (vd - ue)t + p_i,$$

donde $0 \leq p_i < |vd - ue|$. Así $p \equiv p_i \pmod{vd - ue}$, luego

$$ugp_i^2 + dp_i + uf \equiv 0 \pmod{vd - ue}.$$

Tenemos así un número finito de valores posibles para p_i, los que cumplen

$$0 \leq p_i < |vd - ue|, \qquad ugp_i^2 + dp_i + uf \equiv 0 \pmod{vd - ue}.$$

En nuestro ejemplo las condiciones son

$$0 \leq p_i < 11, \qquad 9p_i^2 + 5p_i - 15 \equiv 0 \pmod{11}$$

cuyas soluciones son $p_1 = 9$, $p_2 = 10$.

En general, cualquier p de la forma $p = (vd - ue)t + p_i$ cumplirá que existe un y tal que

$$(vd - ue)y = ugp^2 + dp + uf$$

y, expresando $p = ux + vy$ para dicho y, tendremos un x tal que (x, y) cumple la ecuación, y todas las soluciones son de esta forma.

Vamos a ver en qué condiciones existe un x para un p_i y de un t arbitrario. Para ello observamos que

$$\begin{aligned}(vd - ue)y &= ug((vd - ue)t + p_i)^2 + d((vd - ue)t + p_i) + uf\\ &= ug(vd - ue)^2t^2 + (2ugp_i + d)(vd - ue)t + ugp_i^2 + dp_i + uf\end{aligned}$$

y, en definitiva,

$$y = ug(vd - ue)t^2 + (2ugp_i + d)t + \frac{ugp_i^2 + dp_i + uf}{vd - ue},$$

que es un entero. Ahora calculamos x a partir de $ux = p - vy$, de modo que

$$\begin{aligned}ux &= (vd - ue)t + p_i - uvg(vd - ue)t^2 - (2uvgp_i + vd)t - \frac{uvgp_i^2 + dvp_i + uvf}{vd - ue}\\ &= uvg(ue - vd)t^2 - u(2vgp_i + e)t - \frac{uvgp_i^2 + eup_i + uvf}{vd - ue},\end{aligned}$$

luego

$$x = vg(ue - vd)t^2 - (2vgp_i + e)t - \frac{vgp_i^2 + ep_i + vf}{vd - ue},$$

Para que x sea entero, la condición necesaria y suficiente es que

$$vgp_i^2 + ep_i + vf \equiv 0 \pmod{vd - ue}.$$

Así hemos probado el teorema siguiente:

Teorema 12.33 *Dada una ecuación diofántica*

$$ax^2 + bxy + cy^2 + dx + ey + f = 0,$$

con $D = b^2 - 4ac = 0$, *si* $g = (a, c)$, *podemos expresar* $a = gu^2$, $c = gv^2$, $b = 2guv$, *y entonces la soluciones enteras de la ecuación están determinadas por los números* $0 \leq p_i < |vd - ue|$ *que cumplen*

$$ugp_i^2 + dp_i + uf \equiv 0 \text{ (mód } vd - ue),$$

$$vgp_i^2 + ep_i + vf \equiv 0 \text{ (mód } vd - ue).$$

Para cada p_i *en estas condiciones y cada entero arbitrario* t, *las soluciones de la ecuación son*

$$x = vg(ue - vd)t^2 - (2vgp_i + e)t - \frac{vgp_i^2 + ep_i + vf}{vd - ue},$$

$$y = ug(vd - ue)t^2 + (2ugp_i + d)t + \frac{ugp_i^2 + dp_i + uf}{vd - ue}.$$

Ejemplo Vamos a resolver la ecuación diofántica

$$27x^2 - 36xy + 12y^2 + 5x - 7y - 5 = 0.$$

Tenemos que $g = 3$, $u = 3$, $v = -2$, $d = 5$, $e = -7$, con lo que las congruencias son

$$9p^2 + 5p - 15 \equiv 0 \text{ (mód } 11),$$

$$-6p^2 - 7p + 10 \equiv 0 \text{ (mód } 11),$$

y ambas tienen por solución $p_1 = 9$, $p_2 = 10$. Las soluciones correspondientes a $p = 9$ son

$$(x, y) = (66t^2 + 115t + 49, 99t^2 + 167t + 69),$$

mientras que las correspondientes a $p = 10$ son:

$$(x, y) = (66t^2 + 127t + 60, 99t^2 + 185t + 85).$$

Haciendo el cambio $t \mapsto t - 1$ obtenemos expresiones más simples:

$$\left.\begin{array}{l} x = 66t^2 - 17t \\ y = 99t^2 - 31t + 1 \end{array}\right\} \qquad \left.\begin{array}{l} x = 66t^2 - 5t - 1 \\ y = 99t^2 - 13t - 1 \end{array}\right\}$$

La tabla siguiente muestra las soluciones correspondientes a los primeros valores de t:

-3	-2	-1	0	1	2	3
$(645, 985)$	$(298, 459)$	$(83, 131)$	$(0, 1)$	$(49, 69)$	$(230, 335)$	$(543, 799)$
$(608, 929)$	$(273, 421)$	$(70, 111)$	$(-1, -1)$	$(60, 85)$	$(253, 369)$	$(578, 851)$

■

Capítulo XIII

La aritmética ideal

En el capítulo anterior hemos visto cómo resolver cualquier ecuación diofántica definida por una forma cuadrática, pero todavía no sabemos probar algunas conjeturas sobre representación de números por formas cuadráticas, como por qué un número es de la forma x^2+5y^2 si y sólo si lo es su parte libre de cuadrados. El problema análogo para las formas x^2+y^2, x^2+2y^2 o x^2+3y^2 lo hemos resuelto apoyándonos en la factorización única de los anillos de enteros de los cuerpos $\mathbb{Q}(i)$, $\mathbb{Q}\big(\sqrt{-2}\,\big)$ y $\mathbb{Q}\big(\sqrt{-3}\,\big)$, pero sucede que el cuerpo $\mathbb{Q}\big(\sqrt{-5}\,\big)$, es decir, el anillo de enteros $\mathbb{Z}\big[\sqrt{-5}\,\big]$, no tiene factorización única.

En este capítulo veremos que, desde el punto de vista algebraico adecuado, todos los anillos de enteros algebraicos tienen factorización única, y que esto nos permitirá resolver problemas como el que acabamos de recordar. Consideremos de nuevo el fallo en la factorización única que habíamos encontrado en $\mathbb{Z}\big[\sqrt{-5}\,\big]$:

$$6 = 2\cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5}).$$

Ninguno de los factores es primo. Para que hubiera factorización única, de acuerdo con el teorema 9.13, teniendo en cuenta que

$$x^2+5 \equiv (x+1)^2 \text{ (mód 2)}, \qquad x^2+5 \equiv (x+1)(x-1) \text{ (mód 3)},$$

tendrían que existir

$$\mathfrak{p} = (2, 1+\sqrt{-5}), \quad \mathfrak{q} = (3, 1+\sqrt{-5}), \quad \mathfrak{r} = (3, -1+\sqrt{-5}),$$

de modo que

$$2 = \mathfrak{p}^2, \qquad 3 = \mathfrak{q}\mathfrak{r}.$$

En particular, $\sqrt{-5} \equiv 1 \equiv -1$ (mód $\mathfrak{p}$), con lo que

$$1+\sqrt{-5} \equiv 0 \equiv 1-\sqrt{-5} \text{ (mód } \mathfrak{p}),$$

es decir, $\mathfrak{p} \mid \pm 1+\sqrt{-5}$ y claramente $\mathfrak{q} \mid 1+\sqrt{-5}$, $\mathfrak{r} \mid 1-\sqrt{-5}$, con lo que, teniendo en cuenta las normas, tendríamos las factorizaciones

$$1+\sqrt{5} = \mathfrak{p}\mathfrak{q}, \qquad 1-\sqrt{-5} = \mathfrak{p}\mathfrak{r}.$$

Entonces, las dos factorizaciones de 6 que hemos mostrado no contradirían la factorización única, pues sería

$$6 = 2 \cdot 3 = (\mathfrak{p}^2)(\mathfrak{q}\mathfrak{r}) = (\mathfrak{p}\mathfrak{q})(\mathfrak{p}\mathfrak{r}) = (1+\sqrt{-5})(1-\sqrt{-5}).$$

El único problema es que $\mathfrak{p}$, $\mathfrak{q}$, $\mathfrak{r}$ no existen. No hay en $\mathbb{Z}\big[\sqrt{-5}\,\big]$ elementos de norma 2 ni 3.

Sin embargo, uno de los principios más profundos que han aprendido los matemáticos a lo largo de la historia es que el hecho de que algo no exista no es razón para no hablar de ello. De no ser así, los matemáticos nunca habrían hablado de cosas como $\sqrt{5}$, porque los griegos "sabían" que los números irracionales no existen, ni mucho menos de cosas como $\sqrt{-5}$, porque los números "imaginarios" no existen.

Pues bien: del mismo modo que es posible hablar de números irracionales o imaginarios aunque a primera vista "no existan", vamos a ver que también es posible considerar "factores ideales" como $\mathfrak{p}$, $\mathfrak{q}$, $\mathfrak{r}$ en anillos de enteros algebraicos aunque a primera vista "no existan".

Del mismo modo que, esencialmente, el número $\sqrt{-5}$ está definido especificando quién es su cuadrado, también podemos definir los divisores ideales $\mathfrak{p}$, $\mathfrak{q}$, $\mathfrak{r}$ especificando a qué enteros de $\mathbb{Z}\big[\sqrt{-5}\,\big]$ dividen. Por ejemplo, $\mathfrak{p}$ debería ser el inexistente máximo común divisor de 2 y $1+\sqrt{-5}$, pero, ¿quiénes son los múltiplos del máximo común divisor de dos números dados?

Ejercicio: Probar que si A es un dominio euclídeo y α, $\beta \in A$, entonces los múltiplos del máximo común divisor (α, β) son los elementos de la forma $a\alpha + b\beta$, con a, b en A.

Sucede que el divisor ideal $\mathfrak{p}$ debería dividir a los enteros algebraicos de la forma $a2 + b(1+\sqrt{-5})$, donde a y b recorren $\mathbb{Z}\big[\sqrt{-5}\,\big]$. Si queremos que un divisor ideal esté determinado por los números a los que divide, podemos definirlo como el conjunto de tales números, es decir,

$$\mathfrak{p} = \{a2 + b(1+\sqrt{-5}) \mid a, b \in \mathbb{Z}\big[\sqrt{-5}\,\big]\},$$

de modo que $\mathfrak{p} \mid \alpha$ sea equivalente, por definición, a que $\alpha \in \mathfrak{p}$. Pero obviamente no nos basta considerar un caso particular. Debemos preguntarnos qué queremos que sea un "divisor ideal" $\mathfrak{a}$ en un anillo A en general. Y lo que queremos es que respete las propiedades básicas de la divisibilidad "real". En principio nos bastaría con lo siguiente:

1. $\mathfrak{a} \mid 0$

   (Todos los divisores "reales" dividen a 0, luego los "ideales" también deberían.)

2. Si $\mathfrak{a} \mid a$ y $\mathfrak{a} \mid b$, entonces $\mathfrak{a} \mid a+b$.

3. Si $\mathfrak{a} \mid a$, entonces $\mathfrak{a} \mid ba$.

Si unimos esto a la idea de identificar un divisor ideal con los elementos "reales" a los que debe dividir, llegamos a un concepto algebraico general:

13.1 Ideales

Definición 13.1 Un *ideal* en un dominio A es un conjunto $I \subset A$ que cumpla las propiedades siguientes:

1. $0 \in I$.
2. Si $a, b \in I$, entonces $a + b \in I$.
3. Si $a \in I$ y $b \in A$, entonces $ba \in I$.

Si convenimos en escribir $\mathfrak{a} \mid a$ como sinónimo de $a \in \mathfrak{a}$, tenemos que los ideales de un anillo se comportan como "divisores" en el sentido de que imitan hasta cierto punto el comportamiento de los divisores "reales". Veremos que ese "hasta cierto punto" es mucho, sobre todo en los anillos de enteros algebraicos.

Notemos que la propiedad 3) implica en particular que si $a \in I$, entonces $-a \in I$, lo cual, combinado con la propiedad 2), nos da que si $a, b \in I$, entonces también $a - b \in I$.

Si A es un dominio y $a \in A$, entonces el conjunto de los múltiplos de A, es decir,

$$(a) = \{ba \mid b \in A\}$$

es claramente un ideal de A (precisamente porque las propiedades que hemos exigido a los ideales las hemos elegido entre las propiedades elementales que cumplen los conjuntos de múltiplos). Los ideales de esta forma se llaman *ideales principales* de A.

En particular $0 = (0) = \{0\}$ y $1 = (1) = A$ son ideales de A.

Un *dominio de ideales principales* es un dominio en el que todos los ideales son principales.

Ejemplo En el anillo $\mathbb{Z}\big[\sqrt{-5}\,\big]$, el conjunto

$$\mathfrak{p} = \{a2 + b(1 + \sqrt{-5}) \mid a, b \in \mathbb{Z}\big[\sqrt{-5}\,\big]\}$$

es un ejemplo de ideal que no es principal. Es inmediato comprobar que $\mathfrak{p}$ cumple la definición de ideal. Sin embargo, no puede suceder que $\mathfrak{p} = (\alpha)$, para cierto $\alpha \in \mathbb{Z}\big[\sqrt{-5}\,\big]$, pues en tal caso, como $2, 1+\sqrt{-5} \in \mathfrak{p}$, tendría que ser $\alpha \mid 2$, $\alpha \mid 1 + \sqrt{-5}$, luego, tomando normas, $\mathrm{N}(\alpha) \mid 4$, $\mathrm{N}(\alpha) \mid 6$, pero sabemos que en $\mathbb{Z}\big[\sqrt{-5}\,\big]$ no hay elementos de norma 2, así que tiene que ser $\mathrm{N}(\alpha) = 1$, luego α tiene que ser una unidad, pero entonces $1 = \alpha^{-1}\alpha \in \mathfrak{p}$, y esto es imposible. En efecto, un elemento arbitrario de $\mathfrak{p}$ es de la forma

$$(u + v\sqrt{-5})2 + (v + x\sqrt{-5})(1 + \sqrt{-5}).$$

Desarrollando la expresión se comprueba que si un elemento de esta forma es entero racional entonces es par. ∎

Así pues, $\mathfrak{p}$ es un ejemplo de divisor ideal que no se corresponde con ningún elemento real de su anillo, es decir, que los múltiplos de $\mathfrak{p}$ (los elementos de $\mathfrak{p}$) no son los múltiplos de ningún elemento real del anillo. Por el contrario:

Teorema 13.2 *Todo dominio euclídeo es un dominio de ideales principales.*

DEMOSTRACIÓN: Sea A un dominio euclídeo y sea I un ideal en A. Si I sólo contiene al 0, entonces es $I = (0)$ y es un ideal principal. Supongamos, pues, que I tiene elementos no nulos, y sea $a \in I$ un elemento no nulo de norma euclídea mínima. Si $b \in I$, podemos dividir $b = ac + r$, donde $r = 0$ o bien la norma euclídea de r es menor que la de a. Pero, como $a \in I$, también $r = b - ac \in I$ y, como a tiene norma mínima entre los elementos no nulos de I, tiene que ser $r = 0$. Por lo tanto $b = ac \in (a)$. Esto prueba que $I \subset (a)$ y el recíproco es trivial: si $b \in (a)$, entonces existe un c tal que $b = ca$, y $b \in I$ por definición de ideal. Así pues, $I = (a)$ es principal. ∎

Así pues, en dominios euclídeos como $\mathbb{Z}$ todos los ideales son principales, por lo que el concepto de "divisor ideal" no aporta nada. Cualquier conjunto que cumpla la definición de ideal es el conjunto de los múltiplos de un elemento real del anillo. Pese a ello, veremos que algunos aspectos de la aritmética de un anillo se entienden mejor cuando se expresan en términos de ideales, incluso en el caso de los dominios euclídeos.

Por ejemplo, veamos qué sucede si tratamos de identificar cada elemento a de un dominio íntegro A con el ideal principal (a) que determina. La correspondencia $a \mapsto (a)$ no es unívoca, pero es fácil ver que:

En un dominio íntegro, $(a) = (b)$ si y sólo si a y b son asociados.

En efecto, si a y b son asociados, ambos tienen los mismos múltiplos, y esto es lo que afirma la igualdad $(a) = (b)$. Recíprocamente, si $(a) = (b)$, entonces, como $a \in (b)$, existe un u tal que $a = ub$, e igualmente existe un v tal que $b = va$, luego $a = uva$. Notemos que si $a = 0$, necesariamente $b = 0$ y la conclusión es trivial. En caso contrario, la integridad de A nos permite simplificar y llegamos a que $uv = 1$, luego u y v son unidades de A, luego a y b son asociados.

Sabemos que, para cuestiones aritméticas, dos elementos asociados en un dominio íntegro son esencialmente el mismo elemento, y ahora vemos que, al identificarlos con ideales, dos elementos asociados se convierten literalmente en el mismo ideal. Por ejemplo, los ideales de $\mathbb{Z}$ son

$$0 = (0), \quad 1 = (1) = (-1), \quad (2) = (-2), \quad (3) = (-3), \quad (4) = (-4), \quad \ldots$$

de modo que desaparece la distinción aritméticamente irrelevante entre un número entero y su opuesto. En particular:

En un dominio íntegro, $(a) = 0$ si y sólo si $a = 0$ y $(a) = 1$ si y sólo si a es una unidad.

Así pues, en la aritmética ideal, el ideal nulo $0 = (0) = \{0\}$ representa al número 0, mientras que el ideal unitario $1 = (1) = A$ representa a todas las unidades del dominio A. La divisibilidad se corresponde con la inclusión entre ideales:

En un dominio íntegro, $a \mid b$ si y sólo si $(b) \subset (a)$.

En efecto, si $a \mid b$, entonces todo múltiplo de b es múltiplo de a, y eso es lo que expresa la inclusión $(b) \subset (a)$. Recíprocamente, si $(b) \subset (a)$, entonces $b \in (a)$, luego existe un c tal que $b = ac$, luego $a \mid b$.

Congruencias Observamos ahora que es posible definir congruencias módulo divisores ideales exactamente igual que las hemos definido módulo divisores reales:

Definición 13.3 Sea A un dominio y sea I un ideal de A. Dos elementos a y b de A son *congruentes* módulo I, y lo representaremos por $a \equiv b \pmod{I}$, si $a - b \in I$.

Así, en el caso en que $I = (m)$ es un ideal principal tenemos que

$$a \equiv b \pmod{(m)} \quad \text{si y sólo si} \quad a - b \in (m) \quad \text{si y sólo si} \quad m \mid a - b,$$

luego la congruencia módulo (m) es lo mismo que la congruencia módulo m que ya teníamos definida. Las propiedades siguientes son inmediatas:

1. $a \equiv a \pmod{I}$.
   Pues $a - a = 0 \in I$.
2. *Si* $a \equiv b \pmod{I}$, *entonces* $b \equiv a \pmod{I}$.
   Pues si $a \equiv b \pmod{I}$, entonces $a - b \in I$, luego $b - a = (-1)(a - b) \in I$, luego $b \equiv a \pmod{I}$.
3. *Si* $a \equiv b \pmod{I}$ *y* $b \equiv c \pmod{I}$, *entonces* $a \equiv c \pmod{I}$.
   Tenemos que $a - b, b - c \in I$, luego $a - c = a - b + b - c \in I$.
4. *Si* $a \equiv a' \pmod{I}$, $b \equiv b' \pmod{I}$, *entonces* $a + b \equiv a' + b' \pmod{I}$ *y* $ab \equiv a'b' \pmod{I}$.
   Tenemos que $a - a', b - b' \in I$, luego $a + b - a' - b' = a - a' + b' - b' \in I$ y también
   $$ab - a'b' = ab - a'b + a'b - a'b' = (a - a')b + a'(b - b') \in I.$$

A su vez, esto se traduce en que si representamos por $[a]$ la clase de todos los elementos de A congruentes con a módulo I y llamamos A/I al conjunto de todas las clases de congruencia módulo I, podemos sumar y multiplicar los elementos de A/I con las operaciones dadas por

$$[a] + [b] = [a + b], \quad [a][b] = [ab],$$

sin que importe el representante elegido en la clase, igual que sucede con las congruencias módulo elementos reales, y así A/I se convierte en un anillo con $0 = [0]$ y $1 = [1]$ (que será un dominio siempre que $I \neq A$), llamado *anillo cociente* de A módulo I.

Generadores de ideales La definición del ideal $\mathfrak{p}$ de $\mathbb{Z}\big[\sqrt{-5}\,\big]$ que hemos tomado como ejemplo es un caso particular de una situación muy general:

Definición 13.4 Si A es un dominio, definimos el *ideal generado* por unos elementos $a_1,\dots,a_r$ de A como el conjunto

$$(a_1,\dots,a_r) = \{b_1a_1 + \cdots + b_ra_r \mid b_1,\dots,b_r \in A\}.$$

Es inmediato comprobar que ciertamente se trata de un ideal y, más aún, es el menor ideal de A que contiene (divide) a $a_1,\dots,a_r$. Si $I = (a_1,\dots,a_r)$, diremos que $a_1,\dots,a_r$ son un *sistema generador* del ideal I.

En estos términos $\mathfrak{p} = (2, 1+\sqrt{-5})$. Notemos que estamos empleando la misma notación que usamos para representar el máximo común divisor, y esto no es casual:

Teorema 13.5 *Si A es un dominio y unos elementos de A cumplen*

$$(a_1,\dots,a_r) = (d),$$

entonces d es un máximo común divisor de $a_1,\dots,a_r$.

Demostración: Como $a_i \in (a_1,\dots,a_r) = (d)$, tenemos que $d \mid a_i$, es decir, que d es un divisor común de todos los a_i. Si e es un divisor común de todos los a_i, como $d \in (a_1,\dots,a_r)$, se cumple que $d = c_1a_1 + \cdots + c_ra_r$ y entonces $e \mid d$. ∎

Así, por ejemplo, en $\mathbb{Z}$ tenemos que $(15,10) = (5)$, $(15,3) = (1)$.

En otras palabras, cuando un ideal (a,b) es principal, es que se trata del ideal asociado al máximo común divisor de a y b, de modo que si (a,b) no es principal, podemos pensar en (a,b) como un "máximo común divisor ideal" de a y b. Por otra parte, el teorema anterior implica que en un dominio de ideales principales todo par de elementos tiene un máximo común divisor, y además se cumple la relación de Bezout:

Teorema 13.6 *Si a y b son elementos de un dominio de ideales principales A, entonces tienen un máximo común divisor d y existen elementos u, v en A tales que $d = ua + vb$.*

Demostración: Como A es un dominio de ideales principales, el ideal (a,b) es principal, luego existe un d en A tal que $(a,b) = (d)$, y el teorema anterior implica que d es un máximo común divisor de a y b. Como $d \in (a,b)$, por definición existen u y v en A tales que $d = ua + vb$. ∎

Ejercicio: Probar que el ideal $(2,x)$ en $\mathbb{Z}[x]$ no es principal.

Observemos que, en el anillo $\mathbb{Z}\big[\sqrt{-5}\,\big]$, no es lo mismo en principio

$$(2, 1+\sqrt{-5}) = \{\alpha 2 + \beta(1+\sqrt{-5}) \mid \alpha,\beta \in \mathbb{Z}\big[\sqrt{-5}\,\big]\}$$

que

$$\big\langle 2, 1+\sqrt{-5}\big\rangle = \{a2 + b(1+\sqrt{-5}) \mid a,b \in \mathbb{Z}\}.$$

Sin embargo, vamos a ver que en este caso y en muchos otros, se da la igualdad

$$\mathfrak{p} = (2, 1+\sqrt{-5}) = \left\langle 2, 1+\sqrt{-5}\right\rangle.$$

Esto es consecuencia del teorema siguiente, que muestra que los ideales de un orden cuadrático no son en realidad nada que no conozcamos ya:

Teorema 13.7 *Si $\mathcal{O}_m$ es un orden de un cuerpo cuadrático k, los ideales no nulos de $\mathcal{O}_m$ son los módulos completos $\mathfrak{a} \subset \mathcal{O}_m$ cuyo anillo de coeficientes contiene a $\mathcal{O}_m$.*

Demostración: Ciertamente, si $\mathfrak{a} \subset \mathcal{O}_m$ es un módulo completo cuyo anillo de coeficientes contiene a $\mathcal{O}_m$, entonces $\mathfrak{a}$ es un ideal no nulo de $\mathcal{O}_m$. Las dos primeras propiedades de la definición de ideal las cumple por ser un módulo, y la tercera la cumple porque los elementos de $\mathcal{O}_m$ son coeficientes de $\mathfrak{a}$.

Recíprocamente, supongamos que $\mathfrak{a}$ es un ideal no nulo de $\mathcal{O}_m$. Sea $\alpha \in \mathfrak{a}$ no nulo. Entonces $\bar{\alpha} \in \mathcal{O}_m$, luego, por definición de ideal, $\mathrm{N}(\alpha) = \alpha\bar{\alpha} \in \mathfrak{a}$. Así pues, $\mathfrak{a}$ contiene un número entero no nulo a. Cambiando a por $-a$ si es necesario, podemos tomarlo positivo.

Tenemos que $\mathcal{O}_m = \langle 1, m\omega\rangle$. Consideremos todos los elementos de $\mathfrak{a}$ de la forma $u + vm\omega$, con $0 \le u < a$, $0 \le v < a$. Son un número finito, digamos $\alpha_1, \ldots, \alpha_r$. Vamos a probar que $\mathfrak{a} = (a, \alpha_1, \ldots, \alpha_r)$. Trivialmente se cumple la inclusión $(a, \alpha_1, \ldots, \alpha_r) \subset \mathfrak{a}$. Por otro lado, un elemento de $\mathfrak{a}$ será de la forma $\alpha = c + dm\omega$, para ciertos enteros c y d. Podemos dividir $c = c'a + u$, $d = d'a + v$, con $0 \le u, v < a$. Entonces

$$\alpha = c'a + u + d'am\omega + vm\omega,$$

pero $-c'a$, $-d'am\omega \in \mathfrak{a}$, luego $u + vm\omega \in \mathfrak{a}$, luego $u + vm\omega = \alpha_i$, para cierto i, luego

$$\alpha = (c' + d'm\omega)a + \alpha_i \in (a, \alpha_1, \ldots, \alpha_r).$$

Llamando $\alpha_0 = a$, tenemos que los elementos de $\mathfrak{a}$ son los de la forma

$$(u_0 + v_0 m\omega)\alpha_0 + \cdots + (u_r + v_r m\omega)\alpha_r =$$

$$u_0\alpha_0 + v_0 m\omega\alpha_0 + \cdots + u_r\alpha_r + v_r m\omega\alpha_r,$$

donde u_i, v_i son enteros arbitrarios, luego

$$\mathfrak{a} = \left\langle \alpha_0, m\omega\alpha_0, \ldots, \alpha_r, m\omega\alpha_r\right\rangle.$$

Esto prueba que $\mathfrak{a}$ es un módulo. Falta probar que es completo. En caso contrario, existiría un $\alpha = u + vm\omega \in \mathfrak{a}$ tal que $\mathfrak{a} = \langle\alpha\rangle$, es decir, que todos los elementos de $\mathfrak{a}$ serían de la forma $n(u + vm\omega)$, donde n es un entero arbitrario. Ahora bien, hemos visto que $\mathfrak{a}$ contiene enteros no nulos, luego tiene que ser $v = 0$, pero entonces todos los elementos de $\mathfrak{a}$ son enteros, lo cual es absurdo, pues si $a \in \mathfrak{a}$ es entero, entonces $am\omega \in \mathfrak{a}$ no lo es.

De la propia definición de ideal se sigue que todos los elementos de $\mathcal{O}_m$ son coeficientes de $\mathfrak{a}$. ∎

En particular vemos que los ideales de un orden cuadrático $\mathcal{O}_m$ son módulos enteros contenidos en $\mathcal{O}_m$, pero con la precaución de que su anillo de coeficientes no es necesariamente $\mathcal{O}_m$, sino que puede ser un orden mayor.

Merece la pena reformular el teorema 12.16:

Teorema 13.8 *Sea $k = \mathbb{Q}(\sqrt{d})$ un cuerpo cuadrático y sea $\omega = \sqrt{d}$ o bien $\omega = (1+\sqrt{d})/2$ según el resto de d módulo 4. Entonces:*

1. *Si $\mathfrak{a}$ es un módulo completo de la forma $\mathfrak{a} = k\langle a, b+m\omega\rangle \subset \mathcal{O}_m$, con a, b, k enteros racionales, $\mathfrak{a}$ es un ideal de $\mathcal{O}_m$ si y sólo si $a \mid \mathrm{N}(b+m\omega)$.*
2. *Todo ideal $\mathfrak{a} \subset \mathcal{O}_m$ puede expresarse de esta forma.*
3. *El anillo de coeficientes de $\mathfrak{a}$ es exactamente el orden $\mathcal{O}_m$ si y sólo si $(a, b, m, \mathrm{N}(b+m\omega)/a) = 1$, y en tal caso $\mathrm{N}(\mathfrak{a}) = k^2|a|$.*

En particular, si $\mathfrak{a}$ es un ideal no nulo de un orden $\mathcal{O}_m$, acabamos de ver que es de la forma $\mathfrak{a} = \langle\alpha, \beta\rangle$, pero esto implica a su vez que $\mathfrak{a} = (\alpha, \beta)$, pues, como $\alpha, \beta \in \mathfrak{a}$, tenemos que

$$(\alpha, \beta) \subset \mathfrak{a} = \langle\alpha, \beta\rangle \subset (\alpha, \beta),$$

luego se da la igualdad.

Esto se interpreta como que todo divisor ideal no nulo de $\mathcal{O}_m$ puede expresarse como el máximo común divisor ideal de dos elementos reales de $\mathcal{O}_m$.

Ejemplo En el anillo $\mathbb{Z}[\sqrt{-5}]$ podemos considerar los ideales

$$\begin{aligned}\mathfrak{p} &= \langle 2, 1+\sqrt{-5}\rangle = (2, 1+\sqrt{-5}),\\ \mathfrak{q} &= \langle 3, 1+\sqrt{-5}\rangle = (3, 1+\sqrt{-5}),\\ \mathfrak{r} &= \langle 3, -1+\sqrt{-5}\rangle = (3, -1+\sqrt{-5}).\end{aligned}$$

Todos ellos tienen como anillo de coeficientes $\mathbb{Z}[\sqrt{-5}]$, pues es el mayor orden de $\mathbb{Q}(\sqrt{-5})$, luego el teorema anterior nos da que

$$\mathrm{N}(\mathfrak{p}) = 2, \qquad \mathrm{N}(\mathfrak{q}) = 3 = \mathrm{N}(\mathfrak{r}).$$

■

Normas En el teorema y en el ejemplo precedentes hemos usado que el hecho de que los ideales de los órdenes cuadráticos sean módulos enteros hace que tengamos definida su norma y que ésta sea un número entero. Más aún, para los ideales principales de un orden cuadrático $\mathcal{O}$, en virtud de las propiedades que conocemos de la norma de un módulo, tenemos que[1]

$$\mathrm{N}((\alpha)) = \mathrm{N}(\alpha\mathcal{O}) = |\mathrm{N}(\alpha)|\,\mathrm{N}(\mathcal{O}) = |\mathrm{N}(\alpha)|,$$

[1]Notemos que un ideal principal es similar a $\mathcal{O}$, luego su anillo de coeficiente es necesariamente $\mathcal{O}$ y no puede ser un orden mayor.

de modo que al identificar el "divisor real" α con el ideal (α) la norma es la misma salvo por el signo, que en el caso ideal siempre es positivo.

Observemos ahora que

$$3 \equiv 0 \text{ (mód } \mathfrak{r}), \qquad \sqrt{-5} \equiv 1 \text{ (mód } \mathfrak{r}).$$

A su vez, cualquier elemento de $\mathbb{Z}\big[\sqrt{-5}\,\big]$ es de la forma

$$a + b\sqrt{-5} \equiv a + b \equiv 0, 1, 2 \text{ (mód } \mathfrak{r}),$$

por lo que $\mathbb{Z}\big[\sqrt{-5}\,\big]/\mathfrak{r} = \{[0], [1], [2]\}$. No es difícil ver que las tres clases son distintas entre sí, pero esto es consecuencia de un hecho general que nos da la interpretación natural de la norma de un ideal:

Teorema 13.9 *Si $\mathcal{O}$ es un orden cuadrático y $\mathfrak{a}$ es un ideal de $\mathcal{O}$ cuyo anillo de coeficientes es $\mathcal{O}$, entonces $\mathcal{O}/\mathfrak{a}$ tiene $\mathrm{N}(\mathfrak{a})$ elementos.*

Demostración: Según el teorema 12.16, un ideal de $\mathcal{O}$ puede expresarse en la forma $\mathfrak{a} = k\,\langle a, b + m\omega\rangle$ y su norma es $\mathrm{N}(\mathfrak{a}) = k^2|a|$. Podemos suponer que k y $a > 0$. Por otra parte, $\mathcal{O} = \langle 1, m\omega\rangle = \langle 1, b + m\omega\rangle$. Si llamamos $\gamma = 1 + m\omega$, tenemos que una base de $\mathcal{O}$ es $1, \gamma$ y una base de $\mathfrak{a}$ es $ka, k\gamma$.

Así, todo elemento de $\mathcal{O}$ es de la forma $u + v\gamma$, donde u y v son enteros, y, dividiendo $u = kac + r$, $v = kd + s$, con $0 \le r < ka$, $0 \le s < k$, tenemos que

$$u + v\gamma = cka + kd\gamma + r + s\gamma \equiv r + s\gamma \text{ (mód } \mathfrak{a})$$

luego toda clase de $\mathcal{O}/\mathfrak{a}$ tiene un representante de la forma $r + s\gamma$, luego hay a lo sumo k^2a clases de restos. Por otro lado, si dos clases de este tipo fueran congruentes módulo $\mathfrak{a}$, es decir,

$$r_1 + s_1\gamma \equiv r_2 + s_2\gamma \text{ (mód } \mathfrak{a}),$$

entonces $(r_1 - r_2) + (s_1 - s_2)\gamma \in \mathfrak{a}$, lo que equivale a que

$$ka \mid r_1 - r_2, \qquad k \mid s_1 - s_2,$$

pero $|r_1 - r_2| < ka$, $|s_1 - s_2| < k$, luego $r_1 = r_2$, $s_1 = s_2$. Esto significa que las k^2a clases de congruencia son distintas dos a dos. ∎

Ejercicio: Consideremos el ideal $\mathfrak{a} = \langle 3, 3\sqrt{5}\rangle$ de $\mathcal{O}_3 \subset \mathbb{Q}(\sqrt{5})$. Comprobar que $\mathrm{N}(\mathfrak{a}) = 9$, pero $\mathcal{O}_3/\mathfrak{a}$ tiene 3 elementos.

Una consecuencia elemental del teorema anterior es la siguiente:

> *Si $\mathcal{O}$ es un orden cuadrático, su único ideal de norma 1 es el ideal $1 = (1) = \mathcal{O}$.*

En efecto, sea $\mathfrak{a}$ un ideal de $\mathcal{O}$ de norma 1, y sea $\mathcal{O}'$ su anillo de coeficientes. Entonces $\mathfrak{a} \subset \mathcal{O} \subset \mathcal{O}'$ y el anillo $\mathcal{O}'/\mathfrak{a}$ tiene un único elemento, es decir, es $\{[0]\}$. Esto equivale a que todo elemento de $\mathcal{O}'$ es congruente con 0 módulo $\mathcal{O}'$ o, lo que es lo mismo, a que $\mathfrak{a} = \mathcal{O}'$, luego también $\mathfrak{a} = \mathcal{O}$. ∎

Producto de ideales Otra consecuencia del teorema 13.7 es que tenemos definido el producto de dos ideales de un orden cuadrático. Conviene observar que la definición de producto de ideales es válida en dominios arbitrarios:

$$IJ = \{\sum_{i=1}^{r} a_i b_i \mid a_i \in I, b_i \in J\}.$$

Este producto "ideal" es coherente con el producto "real" del dominio en el sentido de que claramente se cumple:

$$(a)(b) = (ab).$$

En efecto, un elemento de $(a)(b)$ es de la forma

$$\sum_{i=1}^{r} \alpha_i a \beta_i b = \Big(\sum_{i=1}^{r} \alpha_i \beta_i\Big) ab,$$

luego está en (ab), y un elemento de (ab) es de la forma $\alpha ab = (\alpha a)b$, que claramente está en $(a)(b)$.

Notemos que, de las propias definiciones de ideal y de producto, se sigue la inclusión $IJ \subset I \cap J$.

El comportamiento del producto de ideales es especialmente natural en el caso de los ideales de órdenes cuadráticos:

Teorema 13.10 *Si $\mathfrak{a}$, $\mathfrak{b}$ son ideales no nulos de un orden cuadrático $\mathcal{O}$ con anillo de coeficientes $\mathcal{O}$, entonces $\mathfrak{a} \mid \mathfrak{b}$ en el sentido de que existe un ideal $\mathfrak{c}$ tal que $\mathfrak{b} = \mathfrak{a}\mathfrak{c}$ si y sólo si $\mathfrak{b} \subset \mathfrak{a}$.*

Demostración: Si $\mathfrak{b} = \mathfrak{a}\mathfrak{c}$, como $\mathfrak{c} \subset \mathcal{O}$, por definición de coeficiente se cumple que $ac \in \mathfrak{a}$ para todo $a \in \mathfrak{a}$ y todo $c \in \mathfrak{c} \subset \mathcal{O}$. Como $\mathfrak{b}$ está formado por sumas de productos ac, tenemos que $\mathfrak{b} \subset \mathfrak{a}$.

Recíprocamente, si $\mathfrak{b} \subset \mathfrak{a}$, el módulo $\mathfrak{c} = \mathfrak{a}^{-1}\mathfrak{b}$ tiene anillo de coeficientes $\mathcal{O}$ y cumple $\mathfrak{c} = \mathfrak{a}^{-1}\mathfrak{a} \subset \mathfrak{a}^{-1}\mathfrak{a} = \mathcal{O}$, luego es un ideal de $\mathcal{O}$ que trivialmente cumple $\mathfrak{a}\mathfrak{c} = \mathfrak{b}$. ∎

El teorema siguiente es consecuencia inmediata del teorema 12.24:

Teorema 13.11 *Si $\mathfrak{a}$ es un ideal de un orden cuadrático $\mathcal{O}$ con anillo de coeficientes $\mathcal{O}$, entonces $\mathrm{N}(\mathfrak{a}) = \mathfrak{a}\bar{\mathfrak{a}}$.*

Aquí hay que entender que $\mathrm{N}(\mathfrak{a}) = (\mathrm{N}(\mathfrak{a})) = \mathrm{N}(\mathfrak{a})\mathcal{O}$.

Nota También podemos definir una suma de ideales:

$$I + J = \{a + b \mid a \in I,\, b \in J\},$$

es inmediato comprobar que la suma de ideales es un ideal, y también es natural que se le llame "suma", pero conviene tener presente que no generaliza al concepto de suma de "elementos reales" del dominio, sino que en realidad generaliza al concepto de máximo común divisor. En efecto, la suma $I + J$ es el menor ideal tal que $I \subset I + J$ y $J \subset I + J$ y, teniendo en cuenta que una inclusión $I \subset I'$ representa la divisibilidad $I' \mid I$, esto significa que $I + J$ es el menor ideal que divide a I y a J.

En el caso de ideales principales, el teorema 13.5 nos da que

$$(a) + (b) = (c)$$

equivale a que c es el máximo común divisor de a y b (pues claramente se cumple que $(a) + (b) = (a, b)$). ∎

Con esto ya tenemos todos los elementos necesarios para considerar

$$6 = \mathfrak{p}^2\mathfrak{q}\mathfrak{r}$$

como una afirmación con sentido en el anillo $\mathbb{Z}\big[\sqrt{-5}\,\big]$ (donde hay que considerar que 6 denota al ideal principal (6)). El lector puede comprobar que se cumple ciertamente esta igualdad a partir de las definiciones, pero más adelante será inmediata.

Ideales primos Para que podamos considerar la factorización precedente como una descomposición en factores primos necesitamos definir el concepto de ideal primo. La definición es válida en un contexto general:

Definición 13.12 Un ideal P de un dominio A es *primo* si $P \neq A$ y, para todo par de elementos a, b de A, si $ab \in P$, entonces $a \in P$ o $b \in P$.

Nota Ésta es la definición de ideal primo usual en la teoría general de anillos y la más adecuada en muchos contextos, que excluye al ideal $A = (1)$ por definición, pero que permite que 0 pueda ser un ideal primo (y es inmediato que lo es exactamente en el caso de los dominios íntegros). Sin embargo, en el contexto de los anillos de enteros algebraicos es más conveniente excluir al 0, por lo que cuando hablemos de *divisores (ideales)* de un orden cuadrático $\mathcal{O}$ nos referiremos a los ideales (no nulos) de $\mathcal{O}$ y en particular los *divisores primos* de $\mathcal{O}$ serán los ideales primos no nulos de $\mathcal{O}$. ∎

Ejercicio: Probar que, en un dominio íntegro, un elemento p es primo si y sólo si el ideal (p) es primo.

Hay dos caracterizaciones importantes de los ideales primos:

Teorema 13.13 *Si P es un ideal en un dominio A, las condiciones siguientes son equivalentes:*

1. *P es un ideal primo.*
2. *$P \neq A$ y si I, J son ideales de A tales que $IJ \subset P$, entonces $I \subset P$ o $J \subset P$.*
3. *A/P es un dominio íntegro.*

DEMOSTRACIÓN: Si P es primo y se cumple que $IJ \subset P$, pero $I \not\subset P$, podemos tomar $a \in I \setminus P$. Entonces, para todo elemento $b \in J$, se cumple que $ab \in IJ \subset P$, pero $a \notin P$, luego $b \in P$, lo que prueba que $J \subset P$.

Recíprocamente, si P cumple la segunda propiedad y $ab \in P$, entonces $(a)(b) \subset P$, luego $(a) \subset P$ o $(b) \subset P$, luego $a \in P$ o $b \in P$.

Por otra parte, P es primo si y sólo si $A/P \neq \{0\}$ (es decir, si el anillo cociente cumple $1 \neq 0$) y cuando $[a][b] = [0]$, entonces $[a] = 0$ o $[b] = 0$, pero esto es justo la definición de dominio íntegro. ∎

En particular, si $\mathcal{O}$ es el anillo de enteros algebraicos de un cuerpo cuadrático, tenemos que un divisor $\mathfrak{p}$ de $\mathcal{O}$ es primo si y sólo si $\mathfrak{p} \neq 0, 1$ y cuando $\mathfrak{p} \mid \mathfrak{ab}$, entonces $\mathfrak{p} \mid \mathfrak{a}$ o $\mathfrak{p} \mid \mathfrak{b}$.

Más aún, en tal caso $\mathcal{O}/\mathfrak{p}$ no sólo es un dominio íntegro, sino que de hecho es un cuerpo. Esto es consecuencia de que los anillos $\mathcal{O}/\mathfrak{p}$ son finitos, junto con el siguiente hecho elemental:

Teorema 13.14 *Todo dominio íntegro finito es un cuerpo.*

DEMOSTRACIÓN: Si A es un dominio íntegro finito y $a \in A$ no es nulo, entonces las potencias a^n son todas no nulas, pero no pueden ser todas distintas, luego existen números naturales $0 < m < n$ tales que $a^m = a^n$, con lo que $a^m(a^{n-m} - 1) = 0$, luego $a^{n-m} = 1$, luego $a^{n-m-1} \cdot a = 1$, luego a tiene inverso. ∎

Observemos ahora lo siguiente:

> *Si $\mathfrak{a} \neq 1$ es un divisor de un anillo $\mathcal{O}$ de enteros algebraicos de un cuerpo cuadrático, existe un divisor primo $\mathfrak{p}$ de $\mathcal{O}$ tal que $\mathfrak{p} \mid \mathfrak{a}$.*

Basta considerar un divisor $\mathfrak{a} \subset \mathfrak{p} \subsetneq \mathcal{O}$ de norma mínima. Entonces $\mathfrak{p} \mid \mathfrak{a}$, y vamos a probar que es primo probando que $\mathcal{O}/\mathfrak{p}$ es un cuerpo. Si $[a] \in \mathcal{O}/\mathfrak{p}$, entonces $\mathfrak{a} \subset \mathfrak{p} \subset \mathfrak{p} + (a) \subset \mathcal{O}$. Ahora bien, si llamamos $\mathfrak{m} = \mathfrak{p} + (a)$, se cumple[2] que $\mathfrak{m} \mid \mathfrak{p}$. Pongamos que $\mathfrak{p} = \mathfrak{mc}$, luego $\mathrm{N}(\mathfrak{p}) = \mathrm{N}(\mathfrak{m})\,\mathrm{N}(\mathfrak{c})$, y en particular $\mathrm{N}(\mathfrak{m}) \leq \mathrm{N}(\mathfrak{p})$. Por la elección de $\mathfrak{p}$, sólo hay dos posibilidades: o bien $\mathrm{N}(\mathfrak{m}) = \mathrm{N}(\mathfrak{p})$, en cuyo caso $\mathrm{N}(\mathfrak{c}) = 1$, luego $\mathfrak{c} = 1$, luego $\mathfrak{m} = \mathfrak{p}$, luego $a \in \mathfrak{p}$ y $[a] = [0]$, o bien $\mathfrak{m} = 1$, en cuyo caso $\mathfrak{p} + (a) = 1$, luego existen $\pi \in \mathfrak{p}$, $\alpha \in \mathcal{O}$ tales que $\pi + \alpha a = 1$, luego $[\alpha][a] = 1$, luego $[a]$ tiene inverso en $\mathcal{O}/\mathfrak{p}$. ∎

> *Un divisor $\mathfrak{p}$ de un anillo de enteros algebraicos de un cuerpo cuadrático es primo si y sólo si es irreducible en el sentido siguiente: $\mathfrak{p} \neq 0, 1$ y si $\mathfrak{p} = \mathfrak{ab}$, necesariamente $\mathfrak{a} = 1$ o $\mathfrak{b} = 1$.*

[2]Necesiamos la hipótesis de que $\mathcal{O} = \mathcal{O}_1$ para asegurar que el anillo de coeficientes de $\mathfrak{m}$ es también $\mathcal{O}$ y no un orden mayor.

Si $\mathfrak{p}$ es primo y $\mathfrak{p} = \mathfrak{ab}$, entonces $\mathfrak{p} \mid \mathfrak{a}$ o $\mathfrak{p} \mid \mathfrak{b}$. No perdemos generalidad si suponemos el primer caso, y entonces

$$\mathfrak{a} \subset \mathfrak{p} = \mathfrak{ab} \subset \mathfrak{a}.$$

Por lo tanto, $\mathfrak{p} = \mathfrak{pb}$ y esto implica[3] que $\mathfrak{b} = 1$. Recíprocamente, si $\mathfrak{p}$ es irreducible, por la observación precedente existe un primo $\mathfrak{q}$ tal que $\mathfrak{p} = \mathfrak{qa}$, y por la irreducibilidad $\mathfrak{a} = 1$, luego $\mathfrak{p} = \mathfrak{q}$ es primo.

Así pues, la existencia de elementos irreducibles que no son primos que se da en la aritmética "real" de los anillos de enteros algebraicos como $\mathbb{Q}\big(\sqrt{-5}\big)$ no se da en la aritmética ideal.

13.2 Factorización única ideal

Ya estamos en condiciones de probar un resultado fundamental:

Teorema 13.15 *Si $\mathcal{O}$ es el anillo de enteros algebraicos de un cuerpo cuadrático, entonces todo ideal de $\mathcal{O}$ distinto de 0 o 1 se descompone en producto de ideales primos de forma única salvo el orden.*

Demostración: Sea $\mathfrak{a}$ un ideal de $\mathcal{O}$ distinto de 0 o 1. Hemos probado que tiene un divisor primo, de modo que $\mathfrak{a} = \mathfrak{p}_1\mathfrak{a}_1$. Si $\mathfrak{a}_1 \neq 1$, éste tiene a su vez un divisor primo, $\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{a}_2$, de modo que $\mathrm{N}(\mathfrak{a}) > \mathrm{N}(\mathfrak{a}_1) > \mathrm{N}(\mathfrak{a}_2)$. Como las normas no pueden decrecer indefinidamente, al cabo de un número finito de pasos llegamos a un $\mathfrak{a}_r = 1$ y así $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$ es producto de divisores primos.

Supongamos ahora que tenemos dos descomposiciones en primos de un mismo divisor ideal:

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s.$$

Por definición de primo, cada $\mathfrak{p}_i$ divide a un $\mathfrak{q}_j$, y hemos probado que esto implica de hecho que $\mathfrak{p}_i = \mathfrak{q}_j$. Esto significa que en ambos miembros aparecen los mismos divisores primos. Falta probar que cada divisor primo aparece el mismo número de veces. Ahora bien, si un divisor primo apareciera menos veces en un miembro que en el otro, cancelando[4] tales apariciones llegaríamos a dos descomposiciones de un mismo ideal de modo que dicho primo aparecería en una y no en la otra, lo cual ya hemos visto que es imposible. Por lo tanto, ambas descomposiciones son la misma salvo el orden de los factores. ∎

Así pues, vemos que los anillos de enteros algebraicos siempre tienen factorización única ideal, aunque puedan no tener factorización única real. Más aún, sucede que la aritmética ideal es esencialmente la misma que la real si y sólo si hay factorización única real:

[3] Notemos que podemos simplificar $\mathfrak{p}$ porque tiene inverso respecto del producto de módulos.

[4] Recordamos de nuevo que podemos cancelar ideales presentes en los dos miembros multiplicando por el módulo inverso en el grupo de módulos.

Teorema 13.16 *Sea $\mathcal{O}$ el anillo de enteros algebraicos de un cuerpo cuadrático. Entonces $\mathcal{O}$ es un dominio de ideales principales si y sólo si es un dominio de factorización única.*

DEMOSTRACIÓN: Supongamos que $\mathcal{O}$ es un dominio de ideales principales y vamos a probar que tiene factorización única.[5] Para ello tomamos α en $\mathcal{O}$ que no sea nulo ni unitario. Entonces $(\alpha) \neq 0, 1$, luego este ideal admite una descomposición en ideales primos

$$(\alpha) = \mathfrak{p}_1 \cdots \mathfrak{p}_r,$$

pero estos factores primos serán principales:

$$(\alpha) = (\pi_1) \cdots (\pi_r) = (\pi_1 \cdots \pi_r).$$

Como (π_i) es un ideal primo, se cumple que π_i es primo en $\mathcal{O}$, y además la igualdad de ideales equivale a que $\alpha = \epsilon\pi_1 \cdots \pi_r$, donde ϵ es una unidad de $\mathcal{O}$. Cambiando π_1 por $\epsilon\pi_1$ podemos suponer que $\epsilon = 1$, con lo que tenemos α descompuesto en producto de factores primos (en particular irreducibles). Veamos que la descomposición es única.

Ante todo, observemos que el hecho de que en los dominios de ideales principales se cumpla la relación de Bezout implica que en ellos los elementos irreducibles coinciden con los primos.[6] Así, si

$$\alpha = \pi_1 \cdots \pi_r = \rho_1 \cdots \rho_s$$

son dos descomposiciones en factores irreducibles (luego primos), tenemos que

$$(\alpha) = (\pi_1) \cdots (\pi_r) = (\rho_1) \cdots (\rho_s)$$

son dos descomposiciones del ideal (α) en ideales primos, luego, por la unicidad, $r = s$ y, reordenando los factores, podemos suponer que $(\pi_i) = (\rho_i)$, luego cada π_i es asociado a ρ_i, tal y como exige la definición de dominio de factorización única.

Supongamos ahora que $\mathcal{O}$ es un dominio de factorización única. Como el producto de ideales principales es principal, para probar que todos los ideales de $\mathcal{O}$ son principales basta ver que lo es cada divisor primo $\mathfrak{p}$. Sea $\alpha \in \mathfrak{p}$ un elemento no nulo. Notemos que α no puede ser una unidad de $\mathcal{O}$, pues eso implicaría que $\mathfrak{p} = (1)$, en contra de la definición de ideal primo. Por lo tanto, $\alpha = \pi_1 \cdots \pi_r$, para ciertos elementos primos π_i de $\mathcal{O}$. Por lo tanto, $\mathfrak{p} \mid (\alpha) = (\pi_1) \cdots (\pi_r)$, luego existe un i tal que $\mathfrak{p} \mid (\pi_i)$, pero (π_i) es un ideal primo, luego la factorización única ideal implica que $\mathfrak{p} = (\pi_i)$, luego $\mathfrak{p}$ es principal, como había que probar. ∎

[5] En realidad se cumple en general que todo dominio de ideales principales (aunque no sea un anillo de enteros algebraicos) es un dominio de factorización única.

[6] Si π es irreducible y $\pi \mid \alpha\beta$, entonces $(\pi, \alpha) = 1$ o $(\pi, \alpha) = \pi$. En el segundo caso $\pi \mid \alpha$, mientras que en el primero $1 = u\pi + v\alpha$, luego $\beta = u\beta\pi + v\alpha\beta$, luego $\pi \mid \beta$ y por lo tanto π es primo.

Así pues, un anillo de enteros algebraicos sólo contiene ideales no principales cuando éstos hacen falta para "reparar" los fallos en la factorización única real. Si no hay nada que "reparar", los divisores ideales coinciden con los reales y la factorización única ideal es esencialmente la misma que la real.

Ahora podemos "copiar" argumentos sobre la aritmética real de los dominios de factorización única al caso de la aritmética ideal de los anillos de enteros algebraicos. Por ejemplo, si $\mathfrak{p}$ es un divisor primo, tenemos que $\mathfrak{p} \mid \mathrm{N}(\mathfrak{p})$, y descomponiendo la norma en factores primos en $\mathbb{Z}$ concluimos que existe un primo racional p tal que $\mathfrak{p} \mid p$, es decir, que todo divisor primo divide a un primo racional y, más aún, $\mathrm{N}(\mathfrak{p}) \mid \mathrm{N}(p) = p^2$, luego $\mathrm{N}(\mathfrak{p}) = p$ o $\mathrm{N}(\mathfrak{p}) = p^2$. En el segundo caso tiene que ser $\mathfrak{p} = p$, mientras que en el primero $p = \mathrm{N}(\mathfrak{p}) = \mathfrak{p}\bar{\mathfrak{p}}$, si bien cabe la posibilidad de que $\mathfrak{p} = \bar{\mathfrak{p}}$.

Definición 13.17 Sea k un cuerpo cuadrático y $\mathcal{O}$ su anillo de enteros algebraicos. Sea p un primo racional.

1. Se dice que p *se ramifica* en $\mathcal{O}$ (o en k) si su descomposición en factores primos ideales es $p = \mathfrak{p}^2$.

2. Se dice que p *se escinde* en $\mathcal{O}$ si su descomposición en factores primos ideales es $p = \mathfrak{p}_1\mathfrak{p}_2$, donde $\mathfrak{p}_1 \neq \mathfrak{p}_2$.

3. Se dice que p *se conserva* en $\mathcal{O}$ si p es primo en $\mathcal{O}$ (que es equivalente a que (p) sea un ideal primo).

Acabamos de ver que no hay más posibilidades, y ahora podemos probar una versión "ideal" del teorema 9.13 que es válida en todos los anillos de enteros algebraicos. Además, con los resultados que conocemos en estos momentos, la demostración es muy simple:

Teorema 13.18 *Sea $\mathbb{Q}(\sqrt{d})$ un cuerpo cuadrático y $\mathbb{Z}[\omega]$ su anillo de enteros algebraicos. Si $f(x)$ es el polinomio mínimo de ω y p es un primo racional, entonces:*

1. *Si $\bar{f}(x) = (x - \bar{c})^2$ en $\mathbb{Z}_p[x]$, entonces $p = \mathfrak{p}^2$, donde $\mathfrak{p} = (p, \omega - c)$ es un primo cuadrático que cumple $\omega \equiv c \pmod{\mathfrak{p}}$.*

2. *Si $\bar{f}(x) = (x - \bar{c}_1)(x - \bar{c}_2)$ en $\mathbb{Z}_p[x]$, donde $c_1 \not\equiv c_2 \pmod{p}$, entonces $p = \mathfrak{p}_1\mathfrak{p}_2$, donde $\mathfrak{p}_j = (p, \omega - c_j)$ son divisores primos distintos tales que $\omega \equiv c_j \pmod{\mathfrak{p}_j}$.*

3. *Si $\bar{f}(x)$ es irreducible en $\mathbb{Z}_p[x]$, entonces p se conserva primo y ω no es congruente módulo p con ningún entero racional.*

Demostración: Vamos a probar en primer lugar que los ideales primos de norma prima p son precisamente los de la forma $\mathfrak{p} = \langle p, \omega - c\rangle = (p, \omega - c)$, donde $\bar{c}$ es una raíz de $\bar{f}(x)$ en $\mathbb{Z}_p$. Observemos que $f(x) = (x - \omega)(x - \bar{\omega})$.

Por una parte, si $\bar{c}$ es una raíz de $\bar{f}(x)$ en $\mathbb{Z}_p[x]$, entonces

$$\mathrm{N}(\omega - c) = (\omega - c)(\bar{\omega} - c) = f(c) \equiv 0 \pmod{p},$$

luego $p \mid \mathrm{N}(\omega - c)$, y el teorema 13.8 nos da que $\mathfrak{p} = \langle p, \omega - c\rangle$ es un ideal de norma p, luego es primo (no puede descomponerse en factores). Además

$$\mathfrak{p} = \langle p, \omega - c\rangle = (p, \omega - c),$$

pues ciertamente $\mathfrak{p} = \langle p, \omega - c\rangle \subset (p, \omega - c)$ y, como $\mathfrak{p}$ es un ideal que contiene a p y $\omega - c$, también contiene a $(p, \omega - c)$.

Por otra parte, si $\mathfrak{p}$ es un ideal de norma p, sabemos que $\bar{k} = \mathcal{O}/\mathfrak{p}$ es un cuerpo de p elementos que podemos identificar con $\mathbb{Z}_p$. Recordemos los detalles de esta identificación: las clases $[0], \ldots, [p-1]$ son distintas dos a dos, pues se cumple que $a \equiv b \pmod{\mathfrak{p}}$ si y sólo si $\mathfrak{p} \mid b - a$, si y sólo si $\mathrm{N}(\mathfrak{p}) \mid b - a$, si y sólo si $p \mid b - a$, si y sólo si $a \equiv b \pmod{p}$. Por lo tanto, k está formado por estas p clases.

Como $[\omega]$ es una de estas clases, digamos $[c]$, y $\bar{f}([\omega]) = [f(\omega)] = 0$, se cumple que $[f(c)] = 0$, es decir, que $\bar{c}$ es una raíz de $\bar{f}(x)$ en $\mathbb{Z}_p$ y $\omega \equiv c \pmod{\mathfrak{p}}$. Esto significa que $\omega - c \in \mathfrak{p}$, luego $\langle p, \omega - c\rangle \subset \mathfrak{p}$, pero el ideal de la izquierda es primo y múltiplo de $\mathfrak{p}$, luego $\mathfrak{p} = \langle p, \omega - c\rangle$.

Como consecuencia, si $\bar{f}(x) = (x - \bar{c})^2$ en $\mathbb{Z}_p[x]$, el único ideal de norma p es $\mathfrak{p} = \langle p, \omega - c\rangle$, luego tiene que ser $p = \mathfrak{p}^2$. En cambio, si $\bar{f}$ tiene dos raíces distintas c_1 y c_2, los ideales $\mathfrak{p}_j = \langle p, \omega - c_j\rangle$ son dos ideales primos distintos de norma p (pues $\omega \equiv c_j \pmod{\mathfrak{p}_j}$), y ambos tienen que dividir a p, luego $p = \mathfrak{p}_1\mathfrak{p}_2$. Finalmente, si $\bar{f}$ no tiene raíces módulo p, entonces no hay ideales primos de norma p, luego p se conservar primo. ∎

Ahora podemos aplicar la ley de reciprocidad cuadrática a través del teorema 13.18, que nos da:

Teorema 13.19 *Sea k un cuerpo cuadrático, sea $\mathcal{O}$ su anillo de enteros algebraicos y sea χ_k el carácter de k. Entonces un primo p se ramifica, se escinde o se conserva primo en $\mathcal{O}$ si y sólo si $\chi_k(p) = 0, 1, -1$, respectivamente.*

Ejemplo Consideremos de nuevo el ejemplo de factorización no única real

$$30 = 2 \cdot 3 \cdot 5 = \big(1 + \sqrt{-29}\,\big)\big(1 - \sqrt{-29}\,\big)$$

en el anillo $\mathbb{Z}\big[\sqrt{-29}\,\big]$. Puesto que

$$x^2 + 29 \equiv (x+1)^2 \pmod{2}, \qquad x^2 + 29 \equiv (x+1)(x-1) \pmod{3},$$

$$x^2 + 29 \equiv (x+1)(x-1) \pmod{5},$$

tenemos que

$$2 = \mathfrak{p}^2, \qquad 3 = \mathfrak{q}_1\mathfrak{q}_2, \qquad 5 = \mathfrak{r}_1\mathfrak{r}_2,$$

donde

$$\mathfrak{p} = (2, 1+\sqrt{-29}), \quad \mathfrak{q}_1 = (3, 1+\sqrt{-29}), \quad \mathfrak{q}_2 = (3, -1+\sqrt{-29}),$$
$$\mathfrak{r}_1 = (5, 1+\sqrt{-29}), \quad \mathfrak{r}_2 = (5, -1+\sqrt{-29}),$$

luego tenemos la descomposición en factores primos $30 = \mathfrak{p}^2\mathfrak{q}_1\mathfrak{q}_2\mathfrak{r}_1\mathfrak{r}_2$. Por otro lado, es obvio que

$$1+\sqrt{-29} = \mathfrak{p}\mathfrak{q}_1\mathfrak{r}_1, \qquad 1-\sqrt{-29} = \mathfrak{p}\mathfrak{q}_2\mathfrak{r}_2,$$

por lo que las dos descomposiciones de 30 en factores irreducibles reales corresponden a dos agrupaciones distintas de sus factores primos ideales. ∎

Ejemplo Vamos a estudiar la factorización ideal en el anillo de enteros algebraicos $\mathbb{Z}[\sqrt{15}\,]$. Su discriminante es $\Delta = 60$, luego el carácter de $k = \mathbb{Q}(\sqrt{15}\,)$ es una aplicación $\chi_k : U_{60} \longrightarrow \{\pm 1\}$. Sin más que calcular

$$\chi_k(7) = \left(\frac{15}{7}\right) = \left(\frac{1}{7}\right) = 1$$

ya podemos concluir que χ_k viene dado por:

-19	-23	-17	-13	-11	-7	-1	1	7	11	13	17	19	23	29
$-$	$-$	$-$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$-$	$+$	$-$	$-$	$-$

En efecto, teniendo en cuenta que $\chi_k(-1) = 1$ en los cuerpos cuadráticos reales, así como que $\overline{7}^2 = -11$ y $\overline{7}^3 = 17$, obtenemos ocho signos positivos, luego los ocho restantes tienen que ser negativos.

Así pues, en $\mathbb{Z}[\sqrt{15}\,]$ se ramifican los primos $2, 3, 5$ y se escinden los congruentes con $\pm 1, \pm 7, \pm 11, \pm 17$ (mód 60). Los restantes se conservan primos. ∎

Teorema 13.20 *Sea k un cuerpo cuadrático y $\mathcal{O}$ su anillo de enteros algebraicos. El número de ideales de $\mathcal{O}$ de norma n es igual a $\sum_{r|n} \chi_k(r)$.*

Demostración: Descompongamos $n = p_1^{s_1} \cdots p_t^{s_t}$ como producto de factores primos. Teniendo en cuenta la propiedad multiplicativa de χ_k, se cumple que

$$\sum_{r|n} \chi_k(r) = \sum_{i=0}^{s_1} \chi_k(p_1)^i \cdots \sum_{i=0}^{s_t} \chi_k(p_t)^i.$$

Si $\chi_k(p_j) = 0$ entonces $\sum_{i=0}^{s_j} \chi_k(p_j)^i = 1$, luego estos factores no influyen.

Si $\chi_k(p_j) = -1$ entonces $\sum_{i=0}^{s_j} \chi_k(p_j)^i$ vale 1 si s_j es par y 0 si es impar. Por lo tanto la suma total es igual a 0 cuando alguno de los exponentes s_j correspondientes a primos que se conservan es impar. Ciertamente, cuando esto ocurre no hay ideales de norma n.

Si todos estos exponentes son pares entonces el sumatorio se reduce a los factores correspondientes a los primos que se escinden. Supongamos que son $p_1, \ldots, p_a$. Entonces

$$\sum_{r|n} \chi_k(r) = (s_1 + 1) \cdots (s_a + 1).$$

Hay que probar que éste es el número de ideales de norma n. Ahora bien, si $\mathfrak{a}$ es un ideal de norma n y $\mathfrak{p}$ es un primo que divide a un p_j que se ramifica o se conserva, entonces el exponente de $\mathfrak{p}$ en $\mathfrak{a}$ ha de ser $2s_j$ si p_j se ramifica o s_j si p_j se conserva.

La única variación puede darse en los exponentes de los ideales que dividen a primos racionales que se escinden $p_j = \mathfrak{p}\mathfrak{q}$, donde los exponentes de $\mathfrak{p}$ y $\mathfrak{q}$ han de cumplir únicamente que su suma sea s_j. Por lo tanto el exponente de $\mathfrak{p}$ puede ser cualquiera entre 0 y s_j, y éste determina el exponente de $\mathfrak{q}$. Así pues, cada primo p_j que se escinde da lugar a $s_j + 1$ variaciones en la factorización de $\mathfrak{a}$, luego el número de ideales de norma n es $(s_1 + 1) \cdots (s_a + 1)$. ∎

Para acabar de relacionar la aritmética ideal de un anillo de enteros algebraicos con su aritmética real tenemos que considerar un problema que, en realidad, ya hemos resuelto, a saber, el de determinar si un ideal dado es o no principal y, en caso afirmativo, encontrar un generador.

Ejemplo Resolver la ecuación diofántica $x^2 - 10y^2 = 191$ equivale a encontrar los elementos de $\mathbb{Z}\big[\sqrt{10}\,\big]$ de norma 191. Ciertamente, hay ideales con dicha norma, pues

$$x^2 - 10 \equiv (x - 34)(x + 34) \text{ (mód 191)},$$

por lo que $191 = \mathfrak{p}_1\mathfrak{p}_2$, donde

$$\mathfrak{p}_1 = \left\langle 191, 34 + \sqrt{10} \right\rangle, \quad \mathfrak{p}_2 = \left\langle 191, -34 + \sqrt{10} \right\rangle,$$

pero falta decidir si estos ideales son principales y, en tal caso, encontrarles generadores.

Ahora bien, en general, un ideal $\mathfrak{a}$ es principal si es de la forma $(\alpha) = \alpha\mathcal{O}$, es decir, si es similar al módulo $\mathcal{O}$, y sabemos cómo comprobar si se da el caso.

En nuestro ejemplo, queremos determinar si el módulo $\mathfrak{p}_1$ es similar al módulo $\mathbb{Z}\big[\sqrt{10}\,\big] = \langle 1, \sqrt{10}\rangle$, pero esto es justo lo que comprobamos en el ejemplo de la página 438 y la conclusión fue que

$$\mathfrak{p}_1 = (21 - 5\sqrt{10}).$$

Como $\mathfrak{p}_2$ es el conjugado de $\mathfrak{p}_1$, tiene que ser $\mathfrak{p}_2 = (21 + 5\sqrt{10})$.

Ejercicio: Encontrar la expresión general recurrente para las soluciones enteras de la ecuación $x^2 - 10y^2 = 191$.

Ejercicio: Estudiar si los ideales $\mathfrak{q}_1 = \langle 173, 23 + \sqrt{10}\rangle$ y $\mathfrak{q}_2 = \langle 173, -23 + \sqrt{10}\rangle$ son principales y resolver la ecuación diofántica $x^2 - 10y^2 = 173$.

Otra forma de llegar más rápidamente a la misma conclusión es observar que $\mathrm{N}(34+\sqrt{10}) = 191\cdot 6$, luego si $\mathfrak{p}_1$ es principal, su generador será un elemento de norma 191 y $34+\sqrt{10}$ se obtendrá multiplicándolo por un elemento de norma 6.

Ahora bien, es fácil encontrar elementos de norma 6, como $\alpha = 4+\sqrt{10}$. Más precisamente, es fácil ver que 2 se ramifica en $\mathbb{Z}\big[\sqrt{10}\,\big]$, mientras que 3 se escinde, por lo que sólo hay dos ideales de norma 6, que tienen que ser los generados por α y $\bar{\alpha}$, luego si $\mathfrak{p}_1$ es principal, necesariamente $34+\sqrt{10}$ tiene que ser múltiplo de $4\pm\sqrt{10}$, con el signo adecuado. Una simple comprobación muestra que

$$\frac{34+\sqrt{10}}{4+\sqrt{10}} = 21-5\sqrt{10},$$

con lo que hemos encontrado un elemento de norma 191, y es fácil ver que $\mathfrak{p}_1 = (21-5\sqrt{10})$, pues este primo divide a $34+\sqrt{10}$ y sólo un primo de norma 191 cumple esto. ∎

Ejemplo Consideramos ahora la ecuación diofántica $x^2+10y^2 = 659$, cuyas soluciones se corresponden con los elementos de $\mathbb{Z}\big[\sqrt{-10}\,\big]$ de norma 659. Se comprueba que

$$x^2+10 \equiv (x-96)(x+96) \text{ (mód 659)},$$

por lo que hay dos divisores primos de dicha norma:

$$\mathfrak{p}_1 = \left\langle 659, 96+\sqrt{-10}\right\rangle, \qquad \mathfrak{p}_2 = \left\langle 659, -96+\sqrt{-10}\right\rangle.$$

La forma cuadrática asociada a $\mathfrak{p}_1$ es

$$\frac{(x\,659+y(96+\sqrt{-10}))(x\,659+y(96-\sqrt{-10}))}{659} = 659x^2+192xy+14y^2$$

y su reducción es:

$659x^2+192xy+14y^2$	$x=y' \quad y=-x'$
$14x^2-192xy+659y^2$	$x=x'+7y' \quad y=y'$
$14x^2+4xy+y^2$	$x=y' \quad y=-x'$
$x^2-4xy+14y^2$	$x=x'+2y' \quad y=y'$
x^2+10y^2	

Como hemos llegado a la forma asociada al módulo $\mathbb{Z}\big[\sqrt{-10}\,\big]$, concluimos que el ideal es principal. Además, el cambio de variables es

$$\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 7 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 7 \\ -2 & 13 \end{pmatrix}.$$

Por lo tanto, la base de $\mathfrak{p}_1$ que se corresponde con la forma x^2+10y^2 es

$$\begin{pmatrix} -1 & 7 \\ -2 & 13 \end{pmatrix}\begin{pmatrix} 659 \\ 96+\sqrt{-10} \end{pmatrix} = \begin{pmatrix} 13+7\sqrt{-10} \\ -70+13\sqrt{-10} \end{pmatrix}$$

Así hemos encontrado un elemento de $\mathfrak{p}_1$ que cumple $\mathrm{N}(13+7\sqrt{-10}) = 659$, por lo que

$$\mathfrak{p}_1 = (13 + 7\sqrt{-10}), \qquad \mathfrak{p}_2 = (13 - 7\sqrt{-10}).$$

Notemos que no hemos encontrado el generador por casualidad. Lo que tenemos es que

$$\mathfrak{p}_1 = \left\langle 13 + 7\sqrt{-10}, -70 + 13\sqrt{-10} \right\rangle, \qquad \mathbb{Z}\big[\sqrt{-10}\,\big] = \left\langle 1, \sqrt{-10} \right\rangle,$$

donde las dos bases determinan la misma forma cuadrática, y en la prueba del teorema 12.19 hemos visto que en esta situación, necesariamente

$$\mathfrak{p}_1 = \left(\frac{13 + 7\sqrt{-10}}{1}\right) \mathbb{Z}\big[\sqrt{-10}\,\big] = (13 + 7\sqrt{-10}).$$

Por lo tanto, las soluciones de la ecuación $x^2+10y^2 = 659$ son $(x, y) = (\pm 13, \pm 7)$ (las que se obtienen de los generadores que hemos encontrado para los dos divisores primos y las que resultan de multiplicarlos por -1, que es la única unidad de $\mathbb{Z}\big[\sqrt{-10}\,\big]$ distinta de 1).

Ejercicio: Resolver la ecuación diofántica $x^2 + 10y^2 = 173$.

Alternativamente, una vez sabemos que $\mathrm{N}(96 + \sqrt{-10}) = 659 \cdot 14$ y es fácil encontrar elementos de norma 14, por ejemplo $\mathrm{N}(2 \pm \sqrt{-10}) = 14$, y es fácil ver que

$$\frac{96 + \sqrt{-10}}{2 - \sqrt{-10}} = 13 + 7\sqrt{-10},$$

de donde llegamos a la misma conclusión. ∎

En particular, un anillo de enteros algebraicos $\mathcal{O}$ de un cuerpo cuadrático es un dominio de ideales principales si y sólo si todos sus ideales son similares al ideal $\mathcal{O}$, es decir:

Teorema 13.21 *Un anillo de enteros algebraicos de un cuerpo cuadrático es un dominio de ideales principales (lo que equivale a que sea un dominio de factorización única) si y sólo si el número de clases de similitud (no estricta) de dicho anillo es $h = 1$.*

Sabemos cómo comprobar si esto sucede. En particular, ahora es pura rutina comprobar el teorema siguiente:

Teorema 13.22 *Los anillos de enteros algebraicos de los cuerpos $\mathbb{Q}\big(\sqrt{d}\,\big)$ correspondientes a los valores*

$$d = -1, \quad -2, \quad -3, \quad -7, \quad -11, \quad -19, \quad -43, \quad -67 \quad -163.$$

son dominios de factorización única.

Por ejemplo, en el caso de $d = -163$ basta comprobar que la única forma cuadrática reducida de discriminante -163 es $x^2 + xy + 41y^2$.

No es trivial, en cambio, comprobar que son los únicos anillos de enteros algebraicos de cuerpos cuadráticos imaginarios con factorización única. En virtud del teorema 9.10, los cuatro últimos anillos son ejemplos de dominios de ideales principales que no son euclídeos.

Ejemplo En la página 403 hemos visto que hay dos clases de equivalencia estricta de formas cuadráticas de discriminante 56, con representantes

$$x^2 - 14y^2, \qquad 14x^2 - y^2.$$

Por lo tanto, hay dos clases de similitud estricta de módulos con anillo de coeficientes $\mathbb{Z}[\sqrt{14}]$, con representantes $\langle 1, \sqrt{14}\rangle$ y $\langle 14, \sqrt{14}\rangle$. Claramente estos módulos son los ideales (1) y $(\sqrt{14})$, y ambos son principales, luego hay una única clase de similitud no estricta de ideales, lo que prueba que $\mathbb{Z}[\sqrt{14}]$ es un dominio de factorización única. ∎

Ejercicio: Probar que $\mathbb{Z}[\frac{1+\sqrt{53}}{2}]$ es un dominio de factorización única.

Nota Si $\mathcal{O}$ es el anillo de enteros algebraicos de un cuerpo cuadrático k, sabemos que toda clase de similitud de módulos con anillo de coeficientes $\mathcal{O}$ admite un representante entero, es decir, un representante que es un ideal (no nulo) de $\mathcal{O}$. Por ello, si restringimos la relación de similitud a los ideales de $\mathcal{O}$, el grupo de clases de equivalencia resultante es el mismo que si consideramos todos los módulos completos. Por ello a menudo se habla del *grupo de clases de ideales* de un anillo de enteros algebraicos.

Observemos que dos ideales $\mathfrak{a}$ y $\mathfrak{b}$ son similares si y sólo si existe un $\gamma \in k$ no nulo tal que $\mathfrak{b} = \gamma\mathfrak{a}$. No podemos exigir que $\gamma \in \mathcal{O}$, pero, como k es el cuerpo de cocientes de $\mathcal{O}$, sí que podemos expresarlo como $\gamma = \alpha/\beta$, con $\alpha, \beta \in \mathcal{O}$, por lo que la similitud de ideales equivale a que existan $\alpha, \beta \in \mathcal{O}$ tales que $\beta\mathfrak{b} = \alpha\mathfrak{a}$. A su vez, esto equivale a que

$$(\beta)\mathfrak{b} = (\alpha)\mathfrak{a}.$$

En suma, dos ideales son similares si y sólo si pueden convertirse en el mismo ideal multiplicándolos por ideales principales adecuados.

Observemos también que el teorema 12.20 puede reformularse como que todo ideal de $\mathcal{O}$ es similar a uno de norma menor o igual que $\sqrt{|\Delta|/3}$ (o incluso menor o igual que $\sqrt{\Delta/4}$ si $\Delta > 0$), donde Δ es el discriminante de k. ∎

Ejemplo Otra forma de concluir que $\mathbb{Z}[\sqrt{14}]$ tiene factorización única es la siguiente: por la observación precedente, todo ideal de este anillo es similar a uno de norma $\leq \sqrt{14} = 3.7$, es decir, de norma ≤ 3. Como $(14/3) = (2/3) = -1$, el primo 3 se conserva, luego no hay ideales de norma 3 y, como $\mathrm{N}(4+\sqrt{14}) = 2$, vemos que 2 se descompone en producto de ideales primos principales, luego el único ideal de norma 2 es principal (similar a (1)), luego todos los ideales son principales.

Igualmente podemos probar que $\mathbb{Z}[\sqrt{22}]$ es un dominio de ideales principales. En este caso tenemos que todo ideal es similar a uno de norma menor o igual que 4.6, luego, si no es el ideal 1, tiene que ser producto de ideales primos de norma 2 o 3. Ahora bien, es fácil ver que $\mathrm{N}(14+3\sqrt{22}) = -2$ y $\mathrm{N}(5+\sqrt{22}) = 3$, por lo que todos los primos de estas normas son principales, y así todo ideal es principal. ∎

13.3 Aplicaciones

Veamos algunos ejemplos de cómo la factorización única ideal es suficiente, aunque falle la factorización única real, para llevar adelante algunos argumentos aritméticos.

Ejemplo Vamos a resolver la ecuación diofántica $y^2 = x^3 - 13$.

Si y es impar, entonces $y^2 \equiv 1 \pmod 4$, luego $x^3 \equiv 2 \pmod 4$, y es fácil ver que esto es imposible. Por lo tanto, y es par y x es impar. Expresemos la ecuación como

$$(y + \sqrt{-13})(y - \sqrt{-13}) = x^3.$$

Es fácil comprobar que anillo de enteros algebraicos $\mathbb{Z}[\sqrt{-13}]$ tiene número de clases $h = 2$. Esto significa que hay dos clases de ideales, la formada por los ideales principales y otra formada por los ideales no principales.

Observemos que $2 = \mathfrak{q}^2$, donde $\mathfrak{q} = (2, 1 + \sqrt{-13})$ es un ideal no principal, pues tiene norma 2 y, si fuera principal, estaría generado por un elemento de norma 2, y claramente es imposible que $x^2 + 14y^2 = 2$.

Vamos a probar que los dos factores del miembro izquierdo de la ecuación son primos entre sí. Supongamos que un divisor primo $\mathfrak{p}$ los divide a ambos. Entonces divide a su diferencia, es decir, $\mathfrak{p} \mid 2\sqrt{-13}$, pero $\mathfrak{p} \mid x$, luego $\mathfrak{p} \neq \mathfrak{q}$, pues x es impar, luego $\mathfrak{p} \nmid 2$, luego $\mathfrak{p} \mid \sqrt{-13}$, luego $\mathfrak{p} = \sqrt{-13}$, pues este ideal es primo (tiene norma 13). Por lo tanto $\sqrt{-13} \mid x^3$, luego $\sqrt{-13} \mid x$, luego

$$(\sqrt{-13})^3 \mid x^3 = (y + \sqrt{-13})(y - \sqrt{-13}),$$

luego dos de los primos $\sqrt{-13}$ tienen que dividir al mismo factor, es decir, $(\sqrt{-13})^2 \mid y \pm \sqrt{-13}$, lo que equivale a $13 \mid y \pm \sqrt{-13}$, pero esto es imposible, pues los múltiplos de 13 son de la forma $13u + 13v\sqrt{-13}$. Esta contradicción prueba que los dos factores son primos entre sí.

Ahora bien, si el producto de dos divisores ideales primos entre sí es un cubo, ambos deben ser cubos, luego existe un divisor ideal $\mathfrak{a}$ tal que $y + \sqrt{-13} = \mathfrak{a}^3$.

Ahora usamos un argumento crucial: En el grupo de clases se cumple que $[\mathfrak{a}]^3 = 1$ (pues el ideal $\mathfrak{a}^3$ es principal), luego el orden de $[\mathfrak{a}]$ divide a 3, pero también divide al orden del grupo de clases, que es $h = 2$, luego dicho orden es 1, es decir, que $[\mathfrak{a}] = 1$, y esto equivale a que $\mathfrak{a} = (u + v\sqrt{-13})$ es un ideal principal. Así pues:

$$y + \sqrt{-13} = \epsilon(u + v\sqrt{-13})^3,$$

donde ϵ es una unidad, pero las únicas unidades de $\mathbb{Z}\big[\sqrt{-13}\,\big]$ son ± 1 y ambas son cubos, luego no perdemos generalidad si suponemos que $\epsilon = 1$. Así:

$$y + \sqrt{-13} = u(u^2 - 39v^2) + v(3u^2 - 13v^2)\sqrt{-13}.$$

Igualando las coordenadas vemos que $v(3u^2 - 13v^2) = \pm 1$, lo que implica que $v = \pm 1$ y que $3u^2 - 13v^2 = \pm 1$. Así

$$y = u(u^2 - 39), \quad 3u^2 - 13 = \pm 1.$$

Las únicas soluciones de la segunda ecuación son $u = \pm 2$, lo que nos da $y = \pm 70$ y a su vez $x = 17$. Por lo tanto, las únicas soluciones de la ecuación dada son

$$(x, y) = (17, \pm 70).$$

∎

Ejemplo Vamos a probar que la ecuación $y^2 = x^3 - 14$ no tiene soluciones enteras.

Si y es par, entonces $y^2 \equiv 0 \pmod 4$, luego $x^3 \equiv 2 \pmod 4$, lo cual es imposible. Por lo tanto y es impar, y x también. Expresamos la ecuación como

$$(y + \sqrt{-14})(y - \sqrt{-14}) = x^3.$$

Ahora el grupo de clases de $\mathbb{Z}\big[\sqrt{-14}\,\big]$ tiene orden $h = 4$. Observemos que se cumple que $2 = \mathfrak{q}^2$, $7 = \mathfrak{r}^2$, donde

$$\mathfrak{q} = (2, \sqrt{-14}), \qquad \mathfrak{r} = (7, \sqrt{-14}).$$

Si un primo $\mathfrak{p}$ divide a los dos factores del miembro izquierdo de la ecuación, entonces divide a su diferencia, $\mathfrak{p} \mid 2\sqrt{-14} = \mathfrak{q}^3\mathfrak{r}$, pero $\mathfrak{p} \mid x$, que es impar, luego $\mathfrak{p} \neq \mathfrak{q}$ y tiene que ser $\mathfrak{p} = \mathfrak{r}$. Como en el ejemplo anterior concluimos que $\mathfrak{r} \mid x$, luego

$$\mathfrak{r}^3 \mid x^3 = (y + \sqrt{-14})(y + \sqrt{-14}),$$

luego $7 = \mathfrak{r}^2 \mid y \pm \sqrt{-14}$, lo cual es imposible.

Esto prueba que los dos factores del miembro izquierdo de la ecuación son primos entre sí, luego ambos tienen que ser cubos, es decir, que existe un divisor ideal $\mathfrak{a}$ tal que $y + \sqrt{-14} = \mathfrak{a}^3$.

Como en el ejemplo anterior, el orden de $[\mathfrak{a}]$ en el grupo de clases tiene que dividir a 3 y a $h = 4$, luego es 1 y el divisor es principal, digamos $\mathfrak{a} = (u + v\sqrt{-14})$.

A partir de aquí, un desarrollo análogo al del ejemplo anterior lleva a que la ecuación no tiene solución. ∎

Ejercicio: Usar la técnica de los ejemplos anteriores para probar que la ecuación $y^2 = x^3 - 5$ no tiene soluciones enteras. (Véase la página 190.)

Ahora vamos a obtener un curioso teorema de Euler que es consecuencia de la factorización única del anillo de enteros de $\mathbb{Q}\big(\sqrt{-163}\,\big)$:

Teorema 13.23 *El polinomio $f(x) = x^2 + x + 41$ toma valores primos sobre todos los números naturales entre 0 y 39.*

DEMOSTRACIÓN: Es claro que si $m < n$, entonces $p(m) < p(n)$, luego los cuarenta primos que se obtienen son distintos. También es fácil ver que $p(-n-1) = p(n)$, luego los 40 primeros números negativos dan los mismos primos.

Como hemos dicho, la prueba usa la factorización única en $\mathbb{Q}\big(\sqrt{-163}\,\big)$. Un entero de este cuerpo es de la forma $x + y\frac{1+\sqrt{-163}}{2}$, donde x e y son enteros racionales. Si hacemos $y = 1$ obtenemos un entero $\frac{2x+1+\sqrt{-163}}{2}$ con la propiedad de que no tiene divisores enteros racionales no unitarios, y su norma es, precisamente, $x^2 + x + 41$.

Supongamos que $m = x^2 + x + 41$ es compuesto para un cierto entero x tal que $0 \le x \le 39$. Puesto que $m < 40^2 + 40 + 41 = 41^2$, Existirá un primo $p \mid m$ tal que $p \le 39$.

Tenemos p divide al producto de $\frac{2x+1+\sqrt{-163}}{2}$ por su conjugado, pero no puede dividir a ninguno de los dos factores, luego p no es primo en $\mathbb{Q}\big(\sqrt{-163}\,\big)$. Como tiene norma p^2, ha de descomponerse en producto de dos primos de norma p. Digamos que

$$p = \big(s + t\sqrt{-163}\,\big)\,\big(s - t\sqrt{-163}\,\big) = s^2 + 163t^2,$$

donde s y t son ambos enteros o ambos semienteros. Como p es un primo racional, necesariamente $t \neq 0$. Esto nos da la contradicción

$$163 \le (2s)^2 + 163(2t)^2 = 4p \le 4 \cdot 39 = 156.$$

■

Observemos que $p(40) = 40^2 + 40 + 41 = 40 \cdot 41 + 41 = 41^2$ ya no es primo.

Los cuarenta primos alcanzados son:

41, 43, 47, 53, 61, 71, 83, 97, 113, 131, 151, 173, 197, 223, 251,

281, 313, 347, 383, 421, 461, 503, 547, 593, 641, 691, 743, 797, 853,

911, 971, 1 033, 1 097, 1 163, 1 231, 1 301, 1 373, 1 447, 1 523, 1 601.

Ejemplo Ahora ya podemos entender por qué un número es de la forma x^2+5y^2 (resp. de la forma $2x^2+2xy+3y^2$) si y sólo si lo es su parte libre de cuadrados.

Sabemos que las dos formas cuadráticas se corresponden con las dos clases de similitud de módulos de discriminante -20. Como el grupo de clases tiene dos elementos, cada clase es su propia inversa, luego el teorema 12.29 nos da que un número natural m está representado por una de las formas si y sólo si su clase correspondiente contiene un ideal de norma m.

Supongamos que el número es, más concretamente, de la forma k^2m, donde m es libre de cuadrados. Consideremos un ideal $\mathfrak{a}$ de norma k^2m y consideremos su descomposición en factores primos:

$$\mathfrak{a} = q_1^{e_1}\cdots q_r^{e_r}\mathfrak{p}_1^{e'_1}\cdots\mathfrak{p}_s^{e'_s},$$

donde los q_i son los factores primos de norma q_i^2 y los $\mathfrak{p}_i$ son los factores primos de norma prima p_i. Pongamos que $e'_1,\dots,e'_{s'}$ son impares y $e'_{s'+1},\dots,e'_s$ son pares. Sea $k_0 = q_1^{e_1}\cdots q_r^{e_r}$, que tiene norma k_0^2, sean

$$\mathfrak{b} = \mathfrak{p}_1^{(e'_1-1)/2}\cdots\mathfrak{p}_{s'}^{(e'_{s'}-1)/2}\mathfrak{p}_{s'+1}^{e'_{s'+1}/2}\cdots\mathfrak{p}_s^{e'_s/2}, \qquad \mathfrak{c} = \mathfrak{p}_1\cdots\mathfrak{p}_{s'}.$$

Así $\mathfrak{a} = k_0\mathfrak{b}^2\mathfrak{c}$, luego $k^2m = k_0^2\,\mathrm{N}(\mathfrak{b})^2\,\mathrm{N}(\mathfrak{c})$, donde $\mathrm{N}(\mathfrak{c}) = p_1\cdots p_{s'}$ es libre de cuadrados y, por consiguiente, $k = k_0\,\mathrm{N}(\mathfrak{b})$ y $m = \mathrm{N}(\mathfrak{c})$.

Ahora basta observar que, al considerar clases de similitud, $[\mathfrak{a}] = [\mathfrak{b}]^2[\mathfrak{c}] = [\mathfrak{c}]$, pues en un grupo de orden 2, el cuadrado de ambos elementos es el elemento neutro. Concluimos que el ideal $\mathfrak{c}$ de norma m está en la misma clase de similitud que el ideal $\mathfrak{a}$ de norma k^2m, luego k^2m y m están representados por la misma forma.

Por otra parte, ahora tenemos una prueba alternativa conceptualmente más simple de que los números representados por una de las dos formas anteriores son aquellos cuyos primos factores primos con exponente impar cumplen

$$p = 2, 5 \qquad \text{o} \qquad p \equiv 1, 3, 7, 9 \pmod{20}.$$

En efecto, 2 y 5 son los primos que se ramifican en $k = \mathbb{Q}\big(\sqrt{-5}\,\big)$ y los restantes son los que se escinden (los que cumplen $\chi_k(p) = 1$), y es claro que en $\mathbb{Z}\big[\sqrt{-5}\,\big]$ hay ideales de norma m si y sólo si los primos que dividen a m con exponente impar se ramifican o se escinden. (Necesitamos divisores primos de norma p para construir un divisor en cuya norma aparezca p con exponente impar.)

Nos falta determinar, de entre todos estos números, cuáles están representados por una forma y cuáles por la otra. Según hemos visto, basta considerar números libres de cuadrados, y entonces sus factores primos están todos representados por una de las dos formas. Si determinamos por cuál está representado cada primo, en virtud del teorema 12.30 tenemos que el producto estará representado por $2x^2 + 2xy + 3y^2$ si hay un número impar de primos representados por esta forma, y por $x^2 + 5y^2$ en caso contrario.

Notemos que un mismo primo p no puede estar representado por ambas, porque esto significaría que habría divisores primos de norma p en las dos clases de similitud, pero sólo puede haber dos primos de norma p y ambos son conjugados, luego están en clases mutuamente inversas, es decir, ambos están en la misma clase.

Ahora bien, los primos de la forma $p = x^2 + 5y^2$ (distintos de 5) son obviamente restos cuadráticos módulo 5, es decir, que cumplen $p \equiv \pm 1 \pmod 5$, mientras que una comprobación rutinaria muestra que los de la forma $2x^2+2xy+3y^2$ cumplen $p \equiv 2, 3 \pmod 5$, luego ya tenemos probado lo que en su momento habíamos conjeturado:

Los primos de la forma $p = x^2 + 5y^2$ *son los que cumplen*

$$p = 5 \qquad o \qquad p \equiv 1, 9 \text{ (mód 20)},$$

mientras que los de la forma $p = 2x^2+2xy+3y^2$ *son los que cumplen*

$$p = 2 \qquad o \qquad p \equiv 3, 7 \text{ (mód 20)}.$$

Para números libres de cuadrados sólo tenemos que tener en cuenta el teorema 12.30 (véanse las observaciones posteriores a dicho teorema sobre el ejemplo que nos ocupa). Por ejemplo, podemos asegurar que 21 es de la forma $x^2 + 5y^2$, porque 3 y 7 son ambos de la forma $2x^2 + 2xy + 3y^2$. ∎

13.4 Ideales en órdenes no maximales

Resulta natural preguntarse si el teorema 13.15 es válido también para órdenes cuadráticos $\mathcal{O}_m$ con $m > 1$. La respuesta es negativa. Los órdenes cuadráticos no maximales (distintos del anillo de enteros algebraicos de su cuerpo de cocientes) no pueden tener factorización única ideal por la misma razón por la que no pueden tener factorización única real, a saber, porque se cumple la versión ideal del teorema 2.28.

En efecto, sea $\mathcal{O}$ un orden de un cuerpo cuadrático k y supongamos que todo ideal (no nulo) de $\mathcal{O}$ se descompone en forma única en producto de ideales primos. Sea $\alpha \in k$ una raíz de un polinomio mónico

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

con coeficientes enteros. Si el anillo de enteros algebraicos de k es $\mathbb{Z}[\omega]$ y $\mathcal{O} = \mathcal{O}_m$, entonces $\alpha = u + v\omega$, donde u y v son números racionales, y es claro que, si n es el producto de m por los denominadores de u y v, se cumple que $\beta = n\alpha \in \mathcal{O}$, luego $\alpha = \beta/n$.

En particular, hemos expresado $\alpha = r/s$, con $r, s \in \mathcal{O}$. Vamos a probar que $s \mid r$ en $\mathcal{O}$, con lo que tendremos que $\alpha \in \mathcal{O}$, y así será $\mathcal{O} = \mathbb{Z}[\omega]$.

Veamos que si no se cumple $s \mid r$, entonces existe un divisor primo $\mathfrak{p}$ en $\mathcal{O}$ tal que $\mathfrak{p}^{e+1} \mid s$, pero $\mathfrak{p}^{e+1} \nmid r$. En caso contrario, la descomposición en factores primos de s contendría a la de r, luego, multiplicando los ideales sobrantes, obtendríamos un ideal $\mathfrak{a}$ en $\mathcal{O}$ tal que $(s)\mathfrak{a} = (r)$, pero los elementos de $(s)\mathfrak{a}$ son de la forma

$$u_1sa_1 + \cdots u_tsa_t = s(u_1a_1 + \cdots + u_ta_t) = sa,$$

donde $a_i \in \mathfrak{a}$, luego también $a \in \mathfrak{a}$, y así tenemos que $r = sa$, luego $s \mid r$.

Tomando el menor exponente e posible, tenemos que $\mathfrak{p}^e \mid r$. A partir de aquí podemos aplicar literalmente la prueba del teorema 2.28 para llegar a una contradicción.

Pese a esto, vamos a probar que los fallos en la factorización ideal en un orden cuadrático no maximal están muy "localizados". Para ello empezamos probando lo siguiente:

Teorema 13.24 *Sea k un cuerpo cuadrático y $\mathfrak{a}$ un ideal del orden $\mathcal{O}_m$ de k tal que $(\mathrm{N}(\mathfrak{a}), m) = 1$. Entonces el anillo de coeficientes de $\mathfrak{a}$ es $\mathcal{O}_m$.*

DEMOSTRACIÓN: Sea $\mathcal{O}_n$ el anillo de coeficientes de $\mathfrak{a}$, de modo que

$$\mathfrak{a} \subset \mathcal{O}_m \subset \mathcal{O}_n \subset \mathcal{O}_1.$$

Puesto que los elementos de $\mathcal{O}_m$ son los de la forma $a + bm\omega$, con a, b enteros, y los de $\mathcal{O}_m$ son los de la forma $a + bn\omega$, la inclusión requiere que $m = ns$, para cierto número natural s. Basta probar que $s \mid \mathrm{N}(\mathfrak{a})$, pues entonces podemos concluir que $s = 1$, luego $n = m$.

El teorema 13.9 nos da que $\mathcal{O}_n/\mathfrak{a}$ tiene $\mathrm{N}(\mathfrak{a})$ elementos. Cualquiera de ellos es de la forma $[a + bn\omega]$ y, dividiendo $b = ks + r$, con $0 \le r < s$, podemos expresarlo en la forma

$$[a + bn\omega] = [a + km\omega] + [rn\omega],$$

donde el primer sumando es un elemento arbitrario de $\mathcal{O}_m/\mathfrak{a}$.

Además, la expresión es única, pues si

$$[a + km\omega] + [rn\omega] = [a' + k'm\omega] + [r'n\omega],$$

con $0 \le r' < s$, entonces $a - a' + (k - k')m\omega + (r - r')n\omega \in \mathfrak{a} \subset \mathcal{O}_m$, luego también $(r - r')n\omega \in \mathcal{O}_m$, luego $m = sn \mid (r - r')n$, luego $s \mid r - r'$, lo cual sólo es posible si $r = r'$, luego $[a + km\omega] = [a' + k'm\omega]$.

Esto prueba que el número de elementos de $\mathcal{O}_n/\mathfrak{a}$ (es decir, $\mathrm{N}(\mathfrak{a})$) es el producto del número de elementos de $\mathcal{O}_m/\mathfrak{a}$ por s, luego $s \mid \mathrm{N}(\mathfrak{a})$, como queríamos probar.[7] ∎

Ahora podemos probar lo siguiente:

Teorema 13.25 *Sea $\mathcal{O}$ el anillo de enteros algebraicos de un cuerpo cuadrático k y sea $m \ge 1$ un número natural. Entonces la asignación $\mathfrak{a} \mapsto \mathfrak{a} \cap \mathcal{O}_m$ hace corresponder biunívocamente los ideales de $\mathcal{O}$ de norma prima con m con los ideales de $\mathcal{O}_m$ de norma prima con m. Además, esta correspondencia conserva las normas y los productos y $\mathfrak{a}$ es primo si y sólo si lo es $\mathfrak{a} \cap \mathcal{O}_m$.*

DEMOSTRACIÓN: Por el teorema 13.8 sabemos que $\mathfrak{a} = k\langle a, b + \omega\rangle$, donde $a \mid \mathrm{N}(b + m\omega)$, y entonces $\mathrm{N}(\mathfrak{a}) = k^2 a$ (tomando $a > 0$).

Vamos a probar que $\mathfrak{a} \cap \mathcal{O}_m = k\langle a, mb + m\omega\rangle$.

No perdemos generalidad si suponemos $k = 1$, pues, si el resultado es cierto en este caso, en general tenemos que un elemento de $\mathfrak{a} \cap \mathcal{O}_m$ es de la forma $k(u + v\omega)$, con $u + v\omega \in \langle a, b + \omega\rangle$ y como $ku + kv\omega \in \mathcal{O}_m$, tiene que ser $m \mid kv$, luego $m \mid v$, pues $(k, m) = 1$, luego $u + v\omega \in \langle a, b + \omega\rangle \cap \mathcal{O}_m = \langle a, mb + m\omega\rangle$, luego $k(u + v\omega) \in k\langle a, mb + m\omega\rangle$.

[7] Para el lector familiarizado con la teoría de grupos la conclusión es inmediata sin más que observar que $s = |\mathcal{O}_n : \mathcal{O}_m| = |\mathcal{O}_n/\mathfrak{a} : \mathcal{O}_m/\mathfrak{a}| \mid |\mathcal{O}_n/\mathfrak{a}| = \mathrm{N}(\mathfrak{a})$.

Obviamente $\langle a, mb + m\omega\rangle \subset \mathfrak{a} \cap \mathcal{O}_m$. Para probar la inclusión opuesta observamos que los únicos enteros que están en $\mathfrak{a} = \langle a, b + \omega\rangle$ son los múltiplos de a.

En efecto, como $\omega \equiv -b \pmod{\mathfrak{a}}$, todo elemento de $\mathcal{O}$ es congruente con un entero módulo $\mathfrak{a}$ y $a \equiv 0 \pmod{\mathfrak{a}}$. Esto implica que a lo sumo $\mathcal{O}/\mathfrak{a}$ tiene a elementos, las clases de $0, \ldots, a-1$, pero sabemos que tiene exactamente a elementos, luego estos a enteros no son congruentes módulo $\mathfrak{a}$. Si $\mathfrak{a}$ contuviera un entero u que no fuera múltiplo de a, dividiendo $u = ca+r$, obtendríamos otro entero $0 < r < a$ tal que $r \equiv 0 \pmod{\mathfrak{a}}$, y hemos visto que esto es imposible.

Ahora, un elemento de $\mathfrak{a} \cap \mathcal{O}_m$ es e la forma

$$u + vm\omega = u - vmb + v(mb + m\omega).$$

Como el último sumando está en $\mathfrak{a}$, lo mismo vale para el primero, luego, según acabamos de probar, $u - vmb = u'a$, y así

$$u + vm\omega = u'a + v(mb + m\omega) \in \langle a, mb + m\omega\rangle\,.$$

Esto prueba la igualdad. Teniendo en cuenta que $(a, m) = 1$, el teorema 13.8 nos da que el ideal

$$\mathfrak{a} \cap \mathcal{O}_m = k\,\langle a, mb + m\omega\rangle$$

de $\mathcal{O}_m$ tiene anillo de coeficientes $\mathcal{O}_m$ y

$$\mathrm{N}(\mathfrak{a} \cap \mathcal{O}_m) = \mathrm{N}(\mathfrak{a}) = k^2 a.$$

Ahora vamos a probar que si $\mathfrak{a}$ y $\mathfrak{a}'$ son ideales de $\mathcal{O}$ de norma prima con m y $\mathfrak{a} \cap \mathcal{O}_m = \mathfrak{a}' \cap \mathcal{O}_m$, entonces $\mathfrak{a} = \mathfrak{a}'$.

Para ello, basta probar que $\mathfrak{a}$ se puede reconstruir a partir de $\mathfrak{b} = \mathfrak{a} \cap \mathcal{O}_m$ como el conjunto $\mathfrak{a}^*$ de los elementos de $\mathcal{O}$ de la forma[8]

$$\alpha_1\beta_1 + \cdots + \alpha_r\beta_r,$$

con $\alpha_i \in \mathcal{O}$ y $\beta_i \in \mathfrak{b}$. En efecto, es claro que $\mathfrak{a}^* \subset \mathfrak{a}$. Recíprocamente, basta probar que $b + \omega \in \mathfrak{a}^*$, pues entonces lo mismo valdrá para cualquier elemento de $\mathfrak{a}$ (notemos que $a \in \mathfrak{b}$).

Como $(a, m) = 1$, existen enteros u y v tales que $ua + vm = 1$. Así

$$b + \omega = bua + bvm + ua\omega + vm\omega = u(b + \omega)a + v(bm + m\omega) \in \mathfrak{a}^*.$$

Por último, veamos que todo ideal $\mathfrak{a}'$ de $\mathcal{O}_m$ de norma prima con m es de la forma $\mathfrak{a}' = \mathfrak{a} \cap \mathcal{O}_m$, para un cierto ideal $\mathfrak{a}$ de $\mathcal{O}$ de la misma norma.

Por el teorema 13.8 sabemos que $\mathfrak{a}' = k\,\langle a, b' + m\omega\rangle$, de modo que su norma es k^2a, luego $(a, m) = 1$, luego existen enteros tales que $ua + vm = 1$, luego

$$\mathfrak{a}' = k\,\langle a, b'ua + vmb' + m\omega\rangle = k\,\langle a, mvb' + m\omega\rangle = k\,\langle a, mb + m\omega\rangle\,,$$

[8]En particular, observemos que si $\mathfrak{b} = (\beta) = \beta\mathcal{O}_m$, entonces $\alpha^* = (\beta) = \beta\mathcal{O}$. En otras palabras, la correspondencia entre los ideales de $\mathcal{O}$ y los de $\mathcal{O}_m$ hace corresponder los ideales principales generados por elementos de $\mathcal{O}_m$.

luego basta tomar $\mathfrak{a} = k\langle a, b+\omega\rangle$ (que es un ideal, pues tiene que cumplirse que $a \mid \mathrm{N}(mb+m\omega) = m^2\,\mathrm{N}(b+m\omega)$, luego también $a \mid \mathrm{N}(b+\omega)$, ya que a es primo con m).

Con esto hemos probado que la correspondencia es biunívoca y que conserva las normas.

Supongamos ahora que $\mathfrak{a}$ y $\mathfrak{b}$ son dos ideales de $\mathcal{O}$ de norma prima con m que se corresponden con $\mathfrak{a}'$ y $\mathfrak{b}'$ en $\mathcal{O}_m$. Entonces $\mathfrak{a}'\mathfrak{b}'$ también tiene norma prima con m, luego se corresponde con un ideal $(\mathfrak{a}'\mathfrak{b}')^*$ de $\mathcal{O}$. Claramente $\mathfrak{a}'\mathfrak{b}' \subset \mathfrak{a}\mathfrak{b}$, luego $(\mathfrak{a}'\mathfrak{b}')^* \subset \mathfrak{a}\mathfrak{b}$ (por la forma en que hemos construido $(\mathfrak{a}'\mathfrak{b}')^*$), pero ambos ideales tienen la misma norma, luego por la factorización única ideal de $\mathcal{O}$ tiene que ser $(\mathfrak{a}'\mathfrak{b}')^* = \mathfrak{a}\mathfrak{b}$. Equivalentemente, $(\mathfrak{a}\mathfrak{b}) \cap \mathcal{O}_m = (\mathfrak{a} \cap \mathcal{O}_m)(\mathfrak{b} \cap \mathcal{O}_m)$.

Vamos a probar que no sólo es cierto que $\mathfrak{a}$ y $\mathfrak{a}' = \mathfrak{a} \cap \mathcal{O}_m$ tienen la misma norma, sino que, de hecho, los anillos $\mathcal{O}/\mathfrak{a}$ y $\mathcal{O}_m/\mathfrak{a}'$ son esencialmente "el mismo anillo".

Para ello observamos que todo elemento de $\mathcal{O}$ es congruente módulo $\mathfrak{a}$ con un elemento de $\mathcal{O}_m$. En efecto, si la norma de ambos ideales es $N = k^2 a \in \mathfrak{a}$, existen enteros tales que $um + vN = 1$, luego un entero algebraico arbitrario es

$$a + b\omega = a + bum\omega + (vb\omega)N \equiv a + bum\omega \pmod{\mathfrak{a}}.$$

Por otro lado, se cumple

$$a + bm\omega \equiv a' + b'm\omega \pmod{\mathfrak{a}}$$

si y sólo si

$$a + bm\omega \equiv a' + b'm\omega \pmod{\mathfrak{a}'},$$

pues es equivalente que $a + bm\omega - (a' - b'm\omega)$ esté en $\mathfrak{a}$ o en $\mathfrak{a}' = \mathfrak{a} \cap \mathcal{O}_m$, dado que siempre está en $\mathcal{O}_m$.

Esto hace que los representantes de las distintas clases de $\mathcal{O}_m/\mathfrak{a}'$ son también representantes (distintos) de las clases de $\mathcal{O}/\mathfrak{a}$ y, trabajando con representantes en $\mathcal{O}_m$, la suma y el producto de clases se calculan igual en ambos anillos, luego son el mismo anillo.[9]

En particular, $\mathcal{O}/\mathfrak{a}$ es un cuerpo si y sólo si lo es $\mathcal{O}_m/\mathfrak{a}'$, lo que equivale a que $\mathfrak{a}$ es primo si y sólo si lo es $\mathfrak{a}'$. ∎

Ejemplo Consideremos el orden $\mathcal{O}_3 = \langle 1, 3\sqrt{15}\rangle$ del cuerpo cuadrático $\mathbb{Q}(\sqrt{15})$. Como $x^2 - 15 \equiv (x+1)^2 \pmod 2$, en el orden $\mathcal{O}_1$ tenemos que

$$2 = \left\langle 2, 1 + \sqrt{15}\right\rangle^2,$$

luego el teorema anterior nos da que, en $\mathcal{O}_3$, se cumple

$$2 = \left\langle 2, 3 + 3\sqrt{15}\right\rangle^2 = \left\langle 2, 1 + 3\sqrt{15}\right\rangle^2.$$

[9] Técnicamente, son anillos isomorfos.

Similarmente, como $5 = \langle 5, \sqrt{15}\rangle^2$ en $\mathcal{O}_1$, en $\mathcal{O}_3$ tenemos que

$$5 = \left\langle 5, 3\sqrt{15}\right\rangle^2.$$

Igualmente, a partir de $x^2 + 15 \equiv (x+1)(x+6)$ (mód 15) obtenemos la factorización

$$7 = \left\langle 7, 1+\sqrt{15}\right\rangle \left\langle 7, 6+\sqrt{15}\right\rangle$$

en $\mathcal{O}_1$, y a su vez, en $\mathcal{O}_3$,

$$7 = \left\langle 7, 3+3\sqrt{15}\right\rangle \left\langle 7, 18+3\sqrt{15}\right\rangle = \left\langle 7, 3+3\sqrt{15}\right\rangle \left\langle 7, 4+3\sqrt{15}\right\rangle.$$

Ejercicio: Probar que $11 = \langle 11, 6+3\sqrt{15}\rangle \langle 11, 5+3\sqrt{15}\rangle$.

En cambio, el teorema anterior no es aplicable a la descomposición $3 = \mathfrak{p}^2$, donde $\mathfrak{p} = \langle 3, \sqrt{15}\rangle$.

En efecto, observemos que $\mathfrak{p}^* = \mathfrak{p} \cap \mathcal{O}_3 = \langle 3, 3\sqrt{15}\rangle = 3\mathcal{O}_1$, que es un ideal (tanto de $\mathcal{O}_1$ como de $\mathcal{O}_3$) con anillo de coeficientes $\mathcal{O}_1$. En otras palabras, $\mathfrak{p}^*$ es el ideal principal (3) de $\mathcal{O}_1$, de norma 9, pero no hay que confundirlo con el ideal principal (3) del orden $\mathcal{O}_3$, que es $3\mathcal{O}_3 = \langle 3, 9\sqrt{15}\rangle$, cuyo anillo de coeficientes es $\mathcal{O}_3$ y cuya norma también es 9.

Notemos además que $\mathfrak{p}^*$ no es primo como ideal de $\mathcal{O}_1$ (pues $\mathfrak{p}^* \mid (\sqrt{15})^2$, pero $\mathfrak{p}^* \nmid \sqrt{15}$), pero sí que lo es como ideal de $\mathcal{O}_3$, pues si $\alpha\beta \in \mathfrak{p}^*$, para ciertos $\alpha, \beta \in \mathcal{O}_3$, entonces $\alpha\beta \in \mathfrak{p}$, luego $\alpha \in \mathfrak{p}$ o $\beta \in \mathfrak{p}$, luego $\alpha \in \mathfrak{p}^*$ o $\beta \in \mathfrak{p}^*$. En cambio, el ideal principal $(3) = 3\mathcal{O}_3$ de $\mathcal{O}_3$ no es primo, pues $3 \mid (3\sqrt{15})^2$, pero $3 \nmid 3\sqrt{15}$ (en $\mathcal{O}_3$).

Así pues, en $\mathcal{O}_1$ tenemos los ideales $(3) = 3\mathcal{O}_1$ y $\mathfrak{p}$, que son distintos (el primero no es primo, y el segundo sí, el primero tiene norma 9 y el segundo norma 3), pero ambos se corresponden con el mismo ideal primo de $\mathcal{O}_3$ (de norma 9):

$$\mathfrak{p}^* = \mathfrak{p} \cap \mathcal{O}_3 = (3) \cap \mathcal{O}_3,$$

que es distinto del ideal $(3) = 3\mathcal{O}_3$.

Por otra parte: $\mathfrak{p}^{*2} = \langle 9, 135, 9\sqrt{15}\rangle = 3\langle 3, 3\sqrt{15}\rangle = 3\mathfrak{p}^* = 9\mathcal{O}_1$.

En particular, no es cierto que $3 = \mathfrak{p}^{*2}$, ni considerando $3 = 3\mathcal{O}_1$ ni $3 = 3\mathcal{O}_3$. De hecho, el ideal $(3) = 3\mathcal{O}_3$ no es primo, y no admite una descomposición en factores primos ideales, pues los factores tendrían que tener norma 3 y en $\mathcal{O}_3$ no hay ideales de norma 3.

En efecto, un ideal de $\mathcal{O}_3$ de norma 3 tendría que tener anillo de coeficientes $\mathcal{O}_3$ o bien $\mathcal{O}_1$. El segundo caso es imposible, porque el único ideal de $\mathcal{O}_1$ de norma 3 es $\mathfrak{p}$, que no es un ideal de $\mathcal{O}_3$, y el primero tampoco puede darse, como se puede comprobar aplicando el procedimiento explicado en la página 446. Más aún, tenemos la igualdad $\mathfrak{p}^{*2} = (3)\mathfrak{p}^*$ (donde $(3) = 3\mathcal{O}_3$) en la que no podemos simplificar $\mathfrak{p}^*$ para concluir $\mathfrak{p}^* = (3)$.

No obstante, mientras trabajemos con ideales de norma prima con 3, podemos usar la factorización única de $\mathcal{O}_1$ para trabajar en $\mathcal{O}_3$. Por ejemplo, como en $\mathcal{O}_3$ sólo hay dos ideales primos de norma 7 y otros dos de norma 11, concluimos que hay exactamente cuatro ideales de norma 77, que son los cuatro productos posibles de un ideal de norma 7 por otro de norma 11. Equivalentemente, podemos calcular los ideales de norma 77 de $\mathcal{O}_1$ y luego pasar a los correspondientes en $\mathcal{O}_3$. Por ejemplo, uno de ellos es

$$\left\langle 7, 1+\sqrt{15}\right\rangle \left\langle 11, 6+3\sqrt{15}\right\rangle = \left\langle 77, 14+7\sqrt{15}, 11+11\sqrt{15}, 17+3\sqrt{15}\right\rangle$$

$$= \left\langle 77, 57+\sqrt{15}\right\rangle \mapsto \left\langle 77, 171+3\sqrt{15}\right\rangle = \left\langle 77, 17+3\sqrt{15}\right\rangle,$$

que es uno de los cuatro ideales que obtuvimos en el ejemplo de la página 447, e igualmente podemos obtener los otros tres. ∎

Veamos ahora que para calcular grupos de clases de similitud de ideales (estrictas o no estrictas) podemos trabajar únicamente con ideales primos con cualquier entero prefijado. Nos basamos en el hecho siguiente:

Teorema 13.26 *Si $ax^2+bxy+cy^2$ es una forma cuadrática primitiva y m es un entero racional, existe una forma cuadrática $a'x^2+b'xy+c'y^2$ estrictamente equivalente a la dada y tal que $(a', m) = 1$.*

Demostración: Sea r el producto de los primos que dividen a m pero no a c y sea t el producto de los primos que dividen a m y a c pero no a a (se entiende que valen 1 si no hay tales primos).

Entonces $(r, t) = 1$, luego existe un entero u tal que $ur \equiv 1 \pmod t$. Sea finalmente $s = (ur-1)/t$. Así $ru - ts = 1$, luego el cambio de variables $x = rx' + sy'$, $y = tx' + uy'$ transforma la forma cuadrática dada en otra propiamente equivalente en la que $a' = ar^2 + brt + ct^2$.

Veamos que $(a', m) = 1$. Sea p un divisor primo de m. Si $p \nmid c$, entonces $p \mid r$ y $p \nmid t$, luego $p \nmid a'$.

Si $p \mid c$ distinguimos dos casos: si $p \nmid a$ entonces $p \mid t$ pero $p \nmid r$, luego $p \nmid a'$. Si $p \mid c$ y $p \mid a$, como $(a, b, c) = 1$ tenemos que $p \nmid b$, $p \nmid r$ y $p \nmid t$, luego $p \nmid a'$. ∎

Como consecuencia:

Teorema 13.27 *Todo módulo completo M de un cuerpo cuadrático k con anillo de coeficientes $\mathcal{O}_m$ es estrictamente similar a un ideal de $\mathcal{O}_m$ de norma prima con cualquier entero prefijado n.*

Demostración: Consideremos un módulo M cuyo anillo de coeficientes sea $\mathcal{O}_m$. Éste tendrá asociada una forma cuadrática primitiva, y por el teorema anterior, podemos tomarla en la forma $f(x, y) = ax^2 + bxy + cy^2$ con $(a, n) = 1$. Más aún, en la prueba se ve que podemos tomar $a > 0$.

En la prueba del teorema 12.14 se ve que existe un módulo $M' = \langle 1, -\alpha\rangle$ de norma $1/a$ cuya forma asociada es f (y, cambiando α por su conjugado si es

preciso, podemos exigir que la base $1, \alpha$ esté orientada). Por la correspondencia entre módulos y formas, tenemos que M' es estrictamente similar a M.

A su vez, el módulo estrictamente similar $aM' = \langle a, a\gamma\rangle$ tiene norma a. Además, por el teorema 12.9 el anillo de coeficientes de M' es $\mathcal{O}_m = \langle 1, a\gamma\rangle$, luego $aM' \subset \langle a, a\gamma\rangle \subset \mathcal{O}_m$ es un ideal de $\mathcal{O}_m$ de norma a. ∎

Esto implica que, tal y como habíamos afirmado, para trabajar con el grupo de clases de un orden cuadrático $\mathcal{O}_m$ podemos considerar únicamente clases de ideales primos con m, y así aprovechar la factorización única.

Teorema 13.28 *Sea $\mathcal{O}$ el anillo de enteros algebraicos de un cuerpo cuadrático k y sea $m \geq 1$ un número natural. Entonces, si dos ideales del orden $\mathcal{O}_m$ de norma prima con m son (estrictamente) similares, los ideales correspondientes de $\mathcal{O}$ también lo son. Por lo tanto, la correspondencia $I_m(\mathcal{O}_m) \longrightarrow I_m(\mathcal{O})$ dada por el teorema* 13.25 *entre los conjuntos de los ideales de $\mathcal{O}_m$ y de $\mathcal{O}$ de norma prima con m induce una aplicación $\mathcal{H}_m \longrightarrow \mathcal{H}$ entre los grupos de clases de similitud (estricta y no estricta) de $\mathcal{O}_m$ y $\mathcal{O}$.*

Demostración: Sean $\mathfrak{b}$ y $\mathfrak{b}'$ dos ideales de $\mathcal{O}_m$ de norma prima con m y sea γ un elemento no nulo de k (de norma positiva) tal que $\mathfrak{b}' = \gamma\mathfrak{b}$. En la prueba del teorema 13.25 hemos visto que el ideal $\mathfrak{a}$ correspondiente a $\mathfrak{b}$ en $\mathcal{O}$ está formado por los elementos de la forma

$$\alpha_1\beta_1 + \cdots + \alpha_r\beta_r,$$

con $\alpha_i \in \mathcal{O}$ y $\beta_i \in \mathfrak{b}$. Por lo tanto, el ideal correspondiente a $\mathfrak{b}'$ (cuyos elementos son los de la forma $\gamma\beta$, con β en $\mathfrak{b}$), no es sino $\gamma\mathfrak{a}$, luego es (estrictamente) similar a $\mathfrak{a}$.

Por consiguiente, si $[\mathfrak{b}] = [\mathfrak{b}']$ en $\mathcal{H}_m$, también se cumple que $[\mathfrak{a}] = [\mathfrak{a}']$ en $\mathcal{H}_1$, donde $\mathfrak{a}$ y $\mathfrak{a}'$ son los ideales correspondientes a $\mathfrak{b}$ y $\mathfrak{b}'$, respectivamente. ∎

Ejemplo La tabla siguiente muestra la correspondencia dada por el teorema anterior para $k = \mathbb{Q}\big(\sqrt{-14}\,\big)$ y $m = 3$.

La primera columna contiene las formas cuadráticas reducidas de discriminante $D = 3^2 \cdot 4 \cdot (-14) = -504$. La segunda contiene los ideales que se obtienen de ellas de forma natural factorizando las formas, excepto en el caso de la segunda, que nos daría un ideal de norma 3.

$x^2 + 126y^2$ $9x^2 + 14y^2$	$\langle 1, 3\sqrt{-14}\rangle$ $\langle 14, 3\sqrt{-14}\rangle$	$\langle 1, \sqrt{-14}\rangle$ $\langle 14, \sqrt{-14}\rangle$	$x^2 + 14y^2$
$2x^2 + 63y^2$ $7x^2 + 18y^2$	$\langle 2, 3\sqrt{-14}\rangle$ $\langle 7, 3\sqrt{-14}\rangle$	$\langle 2, \sqrt{-14}\rangle$ $\langle 7, \sqrt{-14}\rangle$	$2x^2 + 7y^2$
$5x^2 + 4xy + 26y^2$ $10x^2 - 4xy + 13y^2$	$\langle 5, 2 + 3\sqrt{-14}\rangle$ $\langle 10, -2 + 3\sqrt{-14}\rangle$	$\langle 5, 4 + \sqrt{-14}\rangle$ $\langle 10, 6 + \sqrt{-14}\rangle$	$3x^2 + 2xy + 5y^2$
$5x^2 - 4xy + 26y^2$ $10x^2 + 4xy + 13y^2$	$\langle 5, -2 + 3\sqrt{-14}\rangle$ $\langle 10, 2 + 3\sqrt{-14}\rangle$	$\langle 5, 1 + \sqrt{-14}\rangle$ $\langle 10, 4 + \sqrt{-14}\rangle$	$3x^2 - 2xy + 5y^2$

En efecto, si consideramos, por ejemplo, la séptima forma, factorizando el polinomio $5x^2 - 4x + 26$ obtenemos que

$$5x^2 - 4xy + 26y^2 = 5\left(x + \frac{-2+3\sqrt{-14}}{5}y\right)\left(x + \frac{2-3\sqrt{-14}}{5}y\right),$$

lo que nos lleva al módulo

$$5\left\langle 1, \frac{-2+3\sqrt{-14}}{5}\right\rangle = \left\langle 5, -2+3\sqrt{-14}\right\rangle,$$

que es un ideal de $\mathcal{O}_3$ de norma 5.

Igualmente se procede con todas las demás, salvo en el caso de la segunda, pues si factorizamos el polinomio $9x^2 + 14$ obtenemos:

$$9x^2 + 14y^2 = 9\left(x^2 + \frac{\sqrt{-14}}{3}y\right)\left(x^2 - \frac{\sqrt{-14}}{3}y\right),$$

que nos lleva al ideal de norma 9

$$9\left\langle 1, \frac{\sqrt{-14}}{3}\right\rangle = \left\langle 9, 3\sqrt{-14}\right\rangle$$

y estamos interesados en ideales de norma prima con 3. Por ello pasamos a una forma estrictamente equivalente que no tenga el coeficiente de x^2 múltiplo de 3. La más simple es $14x^2 + 9y^2$, que factoriza como

$$14x^2 + 9y^2 = 14\left(x + \frac{3}{\sqrt{-14}}y\right)\left(x - \frac{3}{\sqrt{-14}}y\right),$$

lo que nos lleva al ideal de norma 14:

$$14\left\langle 1, -\frac{3}{\sqrt{-14}}\right\rangle = \left\langle 14, 3\sqrt{-14}\right\rangle.$$

La tercera columna de la tabla contiene los ideales correspondientes en $\mathcal{O}$. Para calcularlos hay que ajustar el segundo generador para que sea divisible entre 3, por ejemplo, en el caso de la quinta forma tenemos

$$\left\langle 5, 2+3\sqrt{-14}\right\rangle = \left\langle 5, 12+3\sqrt{-14}\right\rangle = \left\langle 5, 3\cdot 4+3\sqrt{-14}\right\rangle,$$

luego le corresponde el ideal $\left\langle 5, 4+\sqrt{-14}\right\rangle$.

Por último, la cuarta columna contiene la forma cuadrática reducida asociada al ideal de la tercera columna. Por ejemplo, el módulo de la última fila tiene asociada la forma $10x^2 + 8xy + 3y^2$, que se reduce a $3x^2 - 2xy + 5y^2$.

Vemos así que el grupo $\mathcal{H}_3$ tiene orden 8, mientras que $\mathcal{H}$ tiene orden 4.

Llamando $\mathfrak{p} = \left\langle 5, 2+3\sqrt{-14}\right\rangle$, se cumple que

$$[\mathfrak{p}]^2 = \left[\left\langle 25, 10+15\sqrt{-14}, -122+12\sqrt{-14}\right\rangle\right] = \left[\left\langle 25, 7+3\sqrt{-14}\right\rangle\right]$$

Al último ideal le corresponde la forma $25x^2 + 14xy + 7y^2$, que se reduce a $7x^2 + 18y^2$, luego $[\mathfrak{p}]^2 = [\langle 7, 3\sqrt{-14}\rangle]$.

Como $\langle 7, 3\sqrt{-14}\rangle^2 = 7$ (pues 7 se ramifica), resulta que $[\mathfrak{p}]^4 = 1$, luego $[\mathfrak{p}]^3 = [\mathfrak{p}]^{-1} = [\langle 5, -2 + 3\sqrt{-14}\rangle]$.

Por otra parte, si llamamos $-1 = [\langle 14, 3\sqrt{-14}\rangle]$, necesariamente

$$-[\mathfrak{p}] = [\langle 10, -2 + 3\sqrt{-14}\rangle], -[\mathfrak{p}]^2 = [\langle 2, 3\sqrt{-14}\rangle], -[\mathfrak{p}]^3 = [\langle 10, 2 + 3\sqrt{-14}\rangle],$$

pues la imagen de -1 en $\mathcal{H}$ es 1, luego la imagen de $-[\mathfrak{p}]$ tiene que ser la misma que la de $[\mathfrak{p}]$, luego tiene que ser la clase que comparte imagen con $[\mathfrak{p}]$, e igualmente en los otros dos casos. En definitiva:

$$\mathcal{H}_3 = \{\pm 1, \pm[\mathfrak{p}], \pm[\mathfrak{p}]^2, \pm[\mathfrak{p}]^3\}$$

y la imagen de cada clase en $\mathcal{H}$ es la que resulta de eliminar los signos y cambiar $[\mathfrak{p}]$ por $[\langle 7, \sqrt{-14}\rangle] = [\langle 2, \sqrt{-14}\rangle]$. ∎

Capítulo XIV

La teoría de los géneros

Toda forma cuadrática de discriminante $D = 60$ es estrictamente equivalente a una de las cuatro formas:[1]

$$x^2 - 15y^2 \qquad 15x^2 - y^2, \qquad 3x^2 - 5y^2, \qquad 5x^2 - 3y^2.$$

La tabla siguiente muestra los primeros números enteros no nulos representados por cada una de ellas:

$x^2 - 15y^2$	-56	-54	-51	-44	-35	-24	-15	-14	-11	-6
	1	4	9	10	16	21	25	34	36	40
$15x^2 - y^2$	-40	-36	-34	-25	-21	-16	-10	-9	-4	-1
	6	11	14	15	24	35	44	51	54	56
$3x^2 - 5y^2$	-45	-42	-33	-32	-20	-18	-17	-8	-5	-2
	3	7	12	22	27	28	30	43	48	55
$5x^2 - 3y^2$	-55	-48	-43	-30	-28	-27	-22	-12	-7	-3
	2	5	8	17	18	20	32	33	42	45

La teoría que conocemos nos permite entender buena parte de la información que contiene esta tabla:

- En primer lugar sabemos que las cuatro clases de equivalencia estrictas determinadas por estas cuatro formas se corresponden con las clases de similitud estricta de módulos (o de ideales) del anillo de enteros algebraicos $\mathbb{Z}\big[\sqrt{15}\,\big]$.
- En segundo lugar tenemos el teorema 12.29, según el cual un número m está representado por una forma asociada a la clase de ideales C si y sólo si la clase C^{-1} contiene un ideal de norma m.
- En tercer lugar, por la factorización única ideal, en $\mathbb{Z}\big[\sqrt{15}\,\big]$ existen ideales de norma m si y sólo si los primos p que dividen a m con exponente impar se ramifican o se escinden en $\mathbb{Z}\big[\sqrt{15}\,\big]$ (de modo que existan ideales primos de norma p).

[1] Véase la página 405.

- Los primos que se ramifican son $p = 2, 3, 5$, y los que se escinden[2] son los que cumplen $p \equiv \pm 1, \pm 7, \pm 11, \pm 17 \pmod{60}$.

- Por lo tanto, un número entero m está representado por una de las cuatro formas si y sólo si los primos que lo dividen con exponente impar están entre los indicados en el punto precedente.

Sin embargo, no tenemos ningún criterio general para determinar cuál (o cuáles) de las cuatro formas representa a un número dado que cumpla las condiciones indicadas. En este capítulo abordaremos el problema, pero antes observemos que todavía sabemos interpretar más a fondo la tabla anterior. Para ello vamos a calcular el grupo de clases de similitud estricta de $\mathbb{Z}\big[\sqrt{15}\,\big]$.

La clase principal $1 = [(1)] = [\langle 1, \sqrt{15}\rangle]$ se corresponde con la forma principal $x^2 - 15y^2$. Por otro lado tenemos la clase de los ideales principales generados por elementos de norma negativa, $-1 = [(\sqrt{15})] = [\langle 15, \sqrt{15}\rangle]$, que se corresponde con la forma $15x^2 - y^2$.

Como $x^2 - 15 \equiv x^2 \pmod 3$, la descomposición en factores de 3 es $3 = \mathfrak{p}^2$, donde $\mathfrak{p} = \langle 3, \sqrt{15}\rangle$, y a este ideal le corresponde claramente la forma $3x^2 - 5y^2$. Similarmente, $5 = \mathfrak{q}^2$, donde $\mathfrak{q} = \langle 5, \sqrt{15}\rangle$ y a este ideal le corresponde la forma $5x^2 - 3y^2$. Ahora bien:

$$-1[\mathfrak{p}] = [\Big\langle 15, \sqrt{15}\Big\rangle \Big\langle 3, \sqrt{15}\Big\rangle] = [\Big\langle 45, 15, 3\sqrt{15}, 15\sqrt{15}\Big\rangle] = [\Big\langle 15, 3\sqrt{15}\Big\rangle]$$

$$= [3\Big\langle 5, \sqrt{15}\Big\rangle] = [\Big\langle 5, \sqrt{15}\Big\rangle] = [\mathfrak{q}].$$

Por lo tanto, podemos representar las clases como ± 1 y $\pm[\mathfrak{p}]$, lo que, combinado con que obviamente $(-1)^2 = [(\sqrt{15})^2] = [(15)] = 1$, nos determina completamente la estructura del grupo de clases:

$[x^2 - 15y^2]$	1
$[15x^2 - y^2]$	-1
$[3x^2 - 5y^2]$	$[\mathfrak{p}]$
$[5x^2 - 3y^2]$	$-[\mathfrak{p}]$

	1	-1	$[\mathfrak{p}]$	$-[\mathfrak{p}]$
1	1	-1	$[\mathfrak{p}]$	$-[\mathfrak{p}]$
-1	-1	1	$-[\mathfrak{p}]$	$[\mathfrak{p}]$
$[\mathfrak{p}]$	$[\mathfrak{p}]$	$-[\mathfrak{p}]$	1	-1
$-[\mathfrak{p}]$	$-[\mathfrak{p}]$	$[\mathfrak{p}]$	-1	1

En particular vemos que cada clase es su propia inversa, por lo que un número m está representado por la forma asociada a una clase C si y sólo si C contiene un ideal de norma m, y esto a su vez se traduce en que si m está representado por la forma de clase C y m' por la forma de clase C', entonces mm' está representado por la forma de clase CC', como se puede observar en la tabla. Por ejemplo, 11 está representado por la forma de clase -1 y 3 por la de clase $[\mathfrak{p}]$, luego 33 está representado por la forma de la clase $-[\mathfrak{p}]$.

Ejercicio: Calcular el ideal $\mathfrak{r}$ que cumple $2 = \mathfrak{r}^2$ e identificar su clase de similitud.

[2] Véase la página 477.

Podemos determinar mediante congruencias qué forma representa cada entero. Para ello observamos que

$$\begin{array}{ll} x^2-15y^2\equiv\ \ x^2\ (\text{mód } 3) & x^2-15y^2\equiv x^2\ \ \ \ (\text{mód } 5)\\ 15x^2-y^2\equiv -y^2\ (\text{mód } 3) & 15x^2-y^2\equiv (2y)^2\ (\text{mód } 5)\\ 3x^2-5y^2\equiv\ \ y^2\ (\text{mód } 3) & 3x^2-5y^2\equiv 3x^2\ \ (\text{mód } 5)\\ 5x^2-3y^2\equiv -x^2\ (\text{mód } 3) & 5x^2-3y^2\equiv 2y^2\ \ (\text{mód } 5) \end{array}$$

Y de aquí se desprende que un entero m no divisible entre 3 ni 5 que esté representado por una forma de discriminante 60, es decir, tal que los primos que lo dividen con exponente impar sean 2 o bien $p\equiv\pm1,\pm7,\pm11,\pm17$ (mód 60) estará representado concretamente por:

$$\begin{array}{ll} x^2-15y^2 & \text{si } m \text{ es un resto cuadrático módulo 3 y 5,}\\ 15x^2-y^2 & \text{si } m \text{ es un resto cuadrático módulo 5, pero no módulo 3,}\\ 3x^2-5y^2 & \text{si } m \text{ es un resto cuadrático módulo 3, pero no módulo 5,}\\ 5x^2-3y^2 & \text{si } m \text{ es no es un resto cuadrático módulo 3 ni módulo 5.} \end{array}$$

Notemos que, en principio, estas condiciones son meramente necesarias, pero como son mutualmente excluyentes y se tiene que dar uno de los cuatro casos, también son suficientes. Así, por ejemplo, podíamos predecir que 22 es de la forma $3x^2-5y^2$ porque sus divisores primos son 2 y otro congruente con 11 módulo 60 (lo que garantiza que puede representarse por una de las cuatro formas) y además

$$22\equiv 1\ (\text{mód } 3),\qquad 22\equiv 2\ (\text{mód } 5),$$

por lo que estamos en el tercero de los cuatro casos anteriores. Para los múltiplos de 3 o 5 podemos usar la estructura del grupo de clases. Por ejemplo, para saber qué forma representa a $795=3\cdot5\cdot53$ observamos en primer lugar que

$$53\equiv -7\ (\text{mód } 60),\qquad 53\equiv 2\ (\text{mód } 3),\qquad 53\equiv 3\ (\text{mód } 5).$$

La primera congruencia nos dice que 53 es representable por una de las cuatro formas, y las otras dos nos dice que es, concretamente, de la forma $5x^2-3y^2$ (de clase $-[\mathfrak{p}]$). Como 5 también es de esta forma, esto implica que $5\cdot53$ es de la forma x^2-15y^2 (porque $(-[\mathfrak{p}])^2=1$) y, como 3 está representado por $3x^2-5y^2$ (de clase $[\mathfrak{p}]$, concluimos que 795 es de la forma $3x^2-5y^2$ (la correspondiente a la clase $1\cdot[\mathfrak{p}]=[\mathfrak{p}]$). En efecto:

$$795=3\cdot30^2-5\cdot9^2.$$

Ejercicio: Probar que un entero $m=3^i\cdot5^j\cdot m^*$ (donde $(m^*,15)=1$) es de la forma x^2-15y^2 si y sólo si 1) los primos que lo dividen con exponente impar son $p=2,3,5$ o bien $p\equiv\pm1,\pm7,\pm11,\pm17$ (mód 60) y 2) se da uno de los cuatro casos siguientes:
2a) i,j son pares y $m^*\equiv1,4$ (mód 15),
2b) i,j son impares y $m^*\equiv-1,-4$ (mód 15),
2c) i es impar, j es par y $m^*\equiv-2,7$ (mód 15),
2d) i es par, j es impar y $m^*\equiv2,-7$ (mód 15).

Ejercicio: Formular un criterio en términos de congruencias que determine qué primos son representables por cada una de las cuatro formas que estamos considerando.

En este capítulo vamos a ver que este ejemplo es un caso particular de la teoría de los géneros de Gauss.

14.1 Equivalencia modular

Vamos a estudiar cuándo un número entero m es representado por una forma cuadrática f módulo otro entero n, es decir, cuándo existen enteros x, y tales que $f(x,y) \equiv m \pmod{n}$. Esto nos dará condiciones necesarias (una para cada n) para que f pueda representar a m, y veremos que en determinadas condiciones de ellas podremos extraer condiciones suficientes.

En el ejemplo que hemos analizado más arriba, nos hemos apoyado en que $x^2 - 15y^2$ representa los cuadrados módulo 3 y módulo 5, mientras que los números representados por las demás formas no estrictamente similares del mismo discriminante (si no son divisibles entre 15) o bien son restos no cuadráticos módulo 3, o bien son restos no cuadráticos módulo 5. Para llegar a esta conclusión nos hemos apoyado en que teníamos representantes de las cuatro clases de forma especialmente simple. En el caso general, necesitaremos aprovechar el hecho de que podemos pasar de unas formas a otras equivalentes más simples, pero respecto a una equivalencia más general que la que estamos considerando hasta ahora:

Definición 14.1 Diremos que dos formas cuadráticas f y g son *equivalentes* módulo un número natural $n > 1$ si existen enteros a, b, c, d tales que

$$f(x,y) \equiv g(ax+by, cx+dy) \pmod{n} \qquad (ad-bc, n) = 1,$$

donde la congruencia hay que entenderla como que los coeficientes de ambas formas son congruentes módulo n.

Observemos que las formas f y g determinan formas $\bar{f}$ y $\bar{g}$ con coeficientes en el anillo $\mathbb{Z}_n$, y el cambio de variables

$$(x,y) = (x',y') \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

determina también un cambio de variables con coeficientes en $\mathbb{Z}_n$ entre $\bar{f}$ y $\bar{g}$. El hecho de que $(ad-bc,n) = 1$ se traduce en que el determinante de la matriz

$$\begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix}$$

(con coeficientes en $\mathbb{Z}_n$) es una unidad de $\mathbb{Z}_n$, por lo que la matriz tiene una matriz inversa, que podemos calcular con la fórmula del teorema 11.7. Si dicha inversa es

$$\begin{pmatrix} \bar{a}' & \bar{c}' \\ \bar{b}' & \bar{d}' \end{pmatrix}$$

entonces la matriz

$$\begin{pmatrix} a' & c' \\ b' & d' \end{pmatrix}$$

determina un cambio de variables tal que

$$g(x,y) \equiv f(a'x + b'y, c'x + d'y) \text{ (mód } n).$$

Equivalentemente, tenemos que

$$\bar{f}(x,y) = \bar{g}(\bar{a}x + \bar{b}y, \bar{c}x + \bar{d}y), \qquad \bar{g}(x,y) = \bar{f}(\bar{a}'x + \bar{b}'y, \bar{c}'x + \bar{d}'y)$$

y en particular, $\bar{f}$ y $\bar{g}$ representan los mismos elementos de $\mathbb{Z}_n$ o, dicho de otro modo, f y g representan los mismos números enteros módulo n.

Ejemplo Consideremos la forma cuadrática

$$g(x,y) = 5x^2 + 9xy - 7y^2.$$

Vamos a estudiarla módulo $n = 7$. Si la reducimos sin más, obtenemos la forma $5x^2 + 2xy$. Si le aplicamos el cambio de variables

$$\left.\begin{array}{rl} x = & x' - 2y' \\ y = & 3y' \end{array}\right\} \quad \text{o, equivalentemente,} \quad (x,y) = (x',y')\begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix}$$

obtenemos la forma $g(x-2y, 3y) = 5x^2 + 7xy - 97y^2$, que módulo 7 se reduce a $f(x,y) = 5x^2 + y^2$, es decir, que tenemos

$$f(x,y) \equiv g(x-2y, 3y) \text{ (mód } 7).$$

El cambio de variables tiene determinante 3, que es primo con 7, luego podemos afirmar que f y g son equivalentes módulo 7. No existe un cambio de variables inverso con coeficientes enteros, pero sí que existe un cambio inverso módulo 7. Para obtenerlo calculamos la matriz inversa (notemos que $1/\bar{3} = \bar{5}$):

$$\begin{pmatrix} \bar{1} & \bar{0} \\ -\bar{2} & \bar{3} \end{pmatrix}^{-1} = \bar{5}\begin{pmatrix} \bar{3} & \bar{0} \\ \bar{2} & \bar{1} \end{pmatrix} = \begin{pmatrix} \bar{1} & \bar{0} \\ \bar{3} & \bar{5} \end{pmatrix}.$$

Esto significa que si aplicamos a f el cambio de variables

$$\left.\begin{array}{rl} x = & x' + 3y' \\ y = & 5y' \end{array}\right\}$$

obtenemos una forma cuadrática $5x^2 + 30xy + 70y^2$, que es congruente con g módulo 7. En otros términos, las formas cuadráticas

$$\bar{g}(x,y) = \bar{5}x^2 + \bar{2}xy, \qquad \bar{f}(x,y) = \bar{5}x^2 + y^2$$

son equivalentes en el sentido de que una se transforma en la otra mediante un cambio de variables con coeficientes en $\mathbb{Z}_7$ que tiene un cambio inverso, también con coeficientes en $\mathbb{Z}_7$.

Por ejemplo, no es inmediato qué elementos de $\mathbb{Z}_7$ están representados por la primera forma, pero sí que lo es que la segunda representa todos los elementos de $\mathbb{Z}_7$, pues $\bar{f}(0,y)$ recorre todos los restos cuadráticos módulo 7 y $\bar{f}(x,0)$ recorre todos los restos no cuadráticos. ∎

En el caso de la equivalencia usual de formas cuadráticas, hemos visto cómo determinar las clases de equivalencia, cuyo número depende esencialmente del discriminante considerado. Sin embargo, vamos a ver que la situación en el caso de la equivalencia modular es mucho más simple. El teorema siguiente nos reduce el problema al caso de la equivalencia módulo potencias de primo:

Teorema 14.2 *Sean m y n dos números naturales primos entre sí. Entonces dos formas cuadráticas son equivalentes módulo mn si y sólo si son equivalentes módulo m y módulo n.*

Demostración: Si tenemos que $f(x,y) \equiv g(a_1x+a_2y, a_3x+a_4y)$ (mód m) y $f(x,y) \equiv g(b_1x+b_2y, b_3x+b_4y)$ (mód n), donde los determinantes de los cambios son primos con m y n respectivamente, por el teorema chino del resto podemos encontrar enteros c_i tales que $c_i \equiv a_i$ (mód m) y $c_i \equiv b_i$ (mód n). Entonces $f(x,y)$ es congruente con $g(c_1x+c_2y, c_3x+c_4y)$ módulo m y módulo n, luego también módulo mn, y es fácil ver que el determinante de este cambio es también primo con mn, luego f y g son equivalentes módulo mn. El recíproco es obvio. ∎

Para estudiar la equivalencia módulo una potencia de primo p^n vamos a buscar formas equivalentes a una dada lo más sencillas posibles. Supongamos primero $p \neq 2$. Dada una forma $f(x,y)$ de discriminante D, el teorema 13.26 nos da otra forma equivalente $ax^2+bxy+cy^2$ tal que $p \nmid a$. El cambio

$$\begin{aligned} x &= x' - by' \\ y &= 2ay' \end{aligned}$$

la transforma en

$$a(x^2 - Dy^2). \tag{14.1}$$

Observemos ahora que la mitad de los elementos de U_{p^n} son cuadrados.

En efecto, observemos en primer lugar que si $\bar{a} \in U_{p^n}$ cumple $\bar{a}^2 = 1$, entonces $\bar{a} = \pm 1$. Basta tener en cuenta que $p^n \mid (a+1)(a-1)$, pero p no puede dividir tanto a $a+1$ como a $a-1$, ya que entonces dividiría a la diferencia, es decir, $p \mid 2$, y p es un primo impar. Así pues, $p^n \mid a \pm 1$, luego $\bar{a} = \pm 1$.

En segundo lugar, si $a^2 = b^2$, entonces $(a/b)^2 = 1$, luego $a/b = \pm 1$, luego $a = \pm b$. Por lo tanto, la aplicación $U_{p^n} \longrightarrow U_{p^n}$ dada por $x \mapsto x^2$ hace corresponde una misma imagen a cada par de elementos $\{a, -a\}$, luego el número de imágenes es la mitad del número de clases de restos en U_{p^n}.

Fijemos un resto no cuadrático cualquiera r módulo p. Obviamente r tampoco es un cuadrado módulo p^n, ni tampoco lo es ninguna clase de la forma $\bar{r}\bar{a}^2$, luego, si $\bar{a}^2$ recorre todos los cuadrados de U_{p^n}, se cumple que $\bar{r}\bar{a}^2$ recorre todos los no cuadrados, luego $U_{p^n} = U_{p^n}^2 \cup rU_{p^n}^2$, donde $U_{p^n}^2$ representa el conjunto de los cuadrados de U_{p^n}.

En particular, el número a de (14.1) se escribirá módulo p^n como $\bar{a} = \bar{u}^2$ o bien $\bar{a} = \bar{r}\bar{u}^2$, para un cierto entero u (primo con p). El cambio $x' = ux$, $y' = uy$ nos transforma (14.1) en una de las dos formas

$$x^2 - Dy^2 \qquad \text{o} \qquad r(x^2 - Dy^2), \tag{14.2}$$

donde —recordemos— r es cualquier resto no cuadrático módulo p que fijemos de antemano. Con esto hemos probado que, si p es impar, hay a lo sumo dos clases de equivalencia de formas cuadráticas de discriminante D módulo p^n.

Supongamos en primer lugar que $p \nmid D$. Es claro que los polinomios x^2 y $r - Dy^2$ toman cada uno $(p+1)/2$ valores distintos módulo p, luego ha de haber enteros u y v que den la misma imagen, es decir, tales que

$$r \equiv u^2 - Dv^2 \ (\text{mód } p).$$

Entonces, $u^2 - Dv^2$ es un resto no cuadrático módulo p, y si elegimos a éste precisamente como r, tenemos la igualdad $r = u^2 - Dv^2$. El cambio de variables

$$\begin{aligned} x &= ux' + Dvy' \\ y &= vx' + uy' \end{aligned}$$

transforma la forma de la izquierda de (14.2) en la forma de la derecha, luego ambas son equivalentes módulo p^n y, en definitiva, todas las formas cuadráticas de discriminante D son equivalentes módulo p^n.

Consideremos ahora el caso en que $p \mid D$. Entonces las formas (14.2) son congruentes módulo p con x^2 y rx^2 respectivamente, que se caracterizan por que una representa sólo restos cuadráticos módulo p y la otra sólo restos no cuadráticos módulo p. En resumen:

Teorema 14.3 *Si p es un primo impar y $p \mid D$, toda forma cuadrática f de discriminante D es equivalente módulo p^n con una de las formas* (14.2) *y sólo con una. Concretamente, f es equivalente a la primera si y sólo si representa restos cuadráticos módulo p y es equivalente a la segunda en caso contrario. Por el contrario, si $p \nmid D$, entonces todas las formas cuadráticas de discriminante D son equivalentes módulo p^n.*

De acuerdo con esto, Gauss dio la definición siguiente:

Definición 14.4 Sea f una forma cuadrática de discriminante D y p un primo impar tal que $p \mid D$. Diremos que f tiene *carácter positivo* módulo p si f representa restos cuadráticos módulo p. En caso contrario se dice que f tiene *carácter negativo* módulo p. Equivalentemente, definimos el *carácter* módulo p de f como

$$\chi_p(f) = \left(\frac{a}{p}\right),$$

donde a es cualquier número representado por f que sea primo con p. Si $p \nmid D$, definimos $\chi_p(f) = 1$.

Las consideraciones anteriores prueban que $\chi_p(f)$ no depende de la elección de a, así como que formas equivalentes módulo p^n tienen el mismo carácter módulo p. En particular, si C es una clase de equivalencia (estricta o no estricta) de formas cuadráticas de discriminante D, podemos definir $\chi_p(C)$ como el carácter de cualquiera de sus miembros. También hemos probado que dos formas f y g de discriminante D son equivalentes módulo p^n si y sólo si tienen el mismo carácter módulo p (notemos que todo esto vale trivialmente cuando $p \nmid D$).

Ejemplo Para las cuatro formas cuadráticas consideradas al principio de este capítulo tenemos que

$$\chi_3(x^2-15y^2) = \chi_3(3x^2-5y^2) = 1, \qquad \chi_3(15x^2-y^2) = \chi_3(5x^2-3y^2) = -1,$$

$$\chi_5(x^2-15y^2) = \chi_5(15x^2-y^2) = 1, \qquad \chi_5(3x^2-5y^2) = \chi_5(5x^2-3y^2) = -1.$$

■

Nos falta estudiar el caso $p = 2$. Si una forma cuadrática tiene discriminante $D = b^2 - 4ac$, entonces D es impar si y sólo si b lo es, y entonces $D \equiv 1 \pmod 8$ si y sólo si $2 \mid ac$. Ocupémonos primero del caso impar, es decir, $p \nmid D$:

Teorema 14.5 *Toda forma cuadrática de discriminante impar D es equivalente módulo 2^n a una de las dos formas*

$$xy \qquad o \qquad x^2 + xy + y^2.$$

Concretamente, una forma es equivalente a la primera si y sólo si $D \equiv 1 \pmod 8$ y es equivalente a la segunda en caso contrario.

Demostración: Por el teorema 13.26, toda forma cuadrática es equivalente a otra $ax^2 + bxy + 2cy^2$ con a impar. Si el discriminante es impar, también lo es b, luego el cambio $y' = by$ nos da otra forma equivalente módulo 2^n en la que $b = 1$. Entonces, la condición $D = 1 - 4ac \equiv 1 \pmod 8$ equivale a que c sea par.

Por otra parte, si aplicamos a xy el cambio de variables $x = x' + 2uy$, $y = ax' + vy$ (con v impar) obtenemos la forma

$$ax^2 + (v + 2au)xy + 2uvy^2.$$

Para que ésta sea congruente con la dada se han de cumplir las congruencias

$$v + 2au \equiv 1 \pmod{2^n} \tag{14.3}$$

$$uv \equiv c \pmod{2^n}. \tag{14.4}$$

Multiplicando por v la primera congruencia y usando la segunda queda

$$v^2 - v \equiv -2ac \pmod{2^n}.$$

Recíprocamente, si demostramos que esta congruencia tiene solución v impar, entonces (14.4) nos permitirá calcular u, y tendremos probado que xy es equivalente módulo 2^n a cualquier forma con discriminante $D \equiv 1 \pmod 8$.

Ahora bien, la congruencia equivale a

$$4v^2 - 4v \equiv -8ac \text{ (mód } 2^{n+2}),$$

o también a

$$(2v-1)^2 \equiv 1 - 8ac \text{ (mód } 2^{n+2}).$$

Según el teorema 7.4, esta congruencia siempre tiene solución v. Además v tiene que ser impar, porque se debe cumplir la misma congruencia módulo 8, y se comprueba entonces que $2v-1$ tiene que ser impar.

Consideremos ahora el caso $D \not\equiv 1 \text{ (mód } 8)$. En primer lugar probamos que si a, b, c y r son impares entonces la congruencia

$$ax^2 + bxy + cy^2 \equiv r \text{ (mód } 2^n) \tag{14.5}$$

tiene solución. Dividiendo entre c podemos suponer $c = 1$. Vamos a encontrar una solución con $y = 1$, de modo que buscamos un x que cumpla

$$ax^2 + bx \equiv r - 1 \text{ (mód } 2^n),$$

que es equivalente a $4a^2x^2 + 4abx \equiv 4a(r-1) \text{ (mód } 2^{n+2}a)$, o también a

$$(2ax + b)^2 \equiv b^2 + 4a(r-1) \text{ (mód } 2^{n+2}a). \tag{14.6}$$

Claramente $b^2 + 4a(r-1) \equiv 1 \text{ (mód } 8)$, luego 7.4 nos da que existe un v tal que

$$v^2 \equiv b^2 + 4a(r-1) \text{ (mód } 2^{n+2}).$$

El teorema chino del resto nos permite tomarlo tal que $v \equiv b \text{ (mód } a)$, y entonces se cumple también $v^2 \equiv b^2 + 4a(r-1) \text{ (mód } a)$, luego

$$v^2 \equiv b^2 + 4a(r-1) \text{ (mód } 2^{n+2}a).$$

Como la congruencia se cumple módulo 8, necesariamente v es impar. Tenemos que $v = au + b$, donde necesariamente $u = 2x$, y así x cumple (14.6).

En particular existen enteros u y v tales que $au^2 + buv + cv^2 \equiv 1 \text{ (mód } 2^n)$. Uno de los dos ha de ser impar. Supongamos que es u. El cambio de variables $x = ux'$, $y = vx' + y$ nos convierte la forma de partida en otra equivalente con $a = 1$. El cambio $y' = by$ nos hace $b = 1$, luego toda forma en el caso que estamos estudiando es equivalente módulo 2^n a una de la forma $x^2 + xy + (2k+1)y^2$.

Vamos a probar que la forma $x^2 + xy + y^2$ se puede transformar en ésta mediante un cambio adecuado. Concretamente hacemos $x = x' + uy'$, $y = vy'$, (con v impar) con lo que llegamos a $x^2 + (2u+v)xy + (u^2 + uv + v^2)y^2$. Hemos de conseguir

$$\begin{aligned} 2u + v &\equiv 1 \text{ (mód } 2^n) \\ u^2 + uv + v^2 &\equiv 2k+1 \text{ (mód } 2^n) \end{aligned}$$

Al despejar v en la primera congruencia y sustituir en la segunda llegamos a la misma congruencia que antes, a saber: $u^2 - u \equiv$ par (mód 2^n), que ya sabemos que tiene solución.

Por último notamos que las dos formas del enunciado no son equivalentes módulo 2^n, pues evidentemente xy representa a todos los enteros, mientras que $x^2 + xy + y^2 \not\equiv 2$ (mód 4). ∎

En particular el teorema anterior prueba que todas las formas cuadráticas con discriminante impar son equivalentes módulo 2^n. Al igual que hemos hecho con los primos impares, definimos el *carácter* módulo 2 de una forma f con discriminante impar como $\chi_2(f) = 1$. Así sigue siendo cierto en este caso que dos formas con el mismo discriminante son equivalentes módulo p^n si y sólo si tienen el mismo carácter módulo p.

Ya sólo nos queda el caso en que 2 divide al discriminante. En tal caso b es par y el teorema 13.26 nos permite suponer que a es impar. Pongamos que la forma es $ax^2 + 2bxy + cy^2$. El cambio $x = x' - by$, $y = ay'$ la transforma en

$$a(x^2 - D'y^2),$$

donde $D' = D/4 = b^2 - ac$.

Sea $r = \pm 1, \pm 5$ de modo que $a \equiv r$ (mód 8). Si $ru \equiv 1$ (mód 2^n), con $n \geq 3$, tenemos que $au \equiv 1$ (mód 8), luego el teorema 7.4 nos da que existe un k (impar) tal que $au \equiv k^2$ (mód 2^n), luego $a \equiv rk^2$ (mód 2^n) y podemos hacer el cambio $x' = kx$, $y' = ky$, con lo que llegamos a una de las cuatro formas:

$$r(x^2 - D'y^2), \qquad r = \pm 1, \pm 5. \tag{14.7}$$

Ahora vamos a ver que si $x^2 - D'y^2$ representa r módulo 8, entonces es equivalente con la correspondiente forma de (14.7) módulo 2^n. En efecto, suponemos que existen enteros u, v tales que $A = u^2 - D'v^2 \equiv r$ (mód 8). El cambio $x = ux' + D'vy'$, $y = vx' + uy'$ nos transforma $x^2 - D'y^2$ en $A(x^2 - D'y^2)$. Ahora expresamos $A \equiv r'k^2$ (mód 2^n), donde $r' = \pm 1, \pm 5$, y al tomar restos módulo 8 queda que $r' = r$, con lo que el cambio $x' = kx$, $y' = ky$ nos lleva a una forma equivalente a (14.7) para el r considerado.

En vista de esto estudiamos los impares representados módulo 8 por la forma $x^2 - D'y^2$. Son los indicados en la tabla siguiente, en función del resto de D' módulo 8:

D'	0	1	2	3	4	5	6	7
r	1	± 1	± 1	1	1	± 1	1	1
		± 5		5	5	± 5	-5	5

La tabla se interpreta como sigue:

- Si $D/4 \equiv 1, 5$ (mód 8) entonces $x^2 - D'y^2$ representa todos los impares módulo 8, luego es equivalente a todas las formas (14.7) y así, todas las formas de discriminante D son equivalentes módulo 2^n.

- Si $D/4 \equiv 3, 4, 7$ (mód 8) entonces las formas $x^2 - D'y^2$ y $5(x^2 - D'y^2)$ son equivalentes, de donde se sigue que $-(x^2 - D'y^2)$ y $-5(x^2 - D'y^2)$ también lo son. Por lo tanto toda forma de discriminante D es equivalente a $\pm(x^2 - D'y^2)$, y estas dos no son equivalentes entre sí, pues una representa sólo los impares congruentes con $1, 5$ módulo 8, y obviamente, la otra sólo representa los congruentes con $-1, -5$ módulo 8.
- Si $D/4 \equiv 2$ (mód 8) tenemos que $\pm(x^2 - D'y^2)$ son equivalentes, y por lo tanto $\pm 5(x^2 - D'y^2)$ también lo son. Toda forma de discriminante D es equivalente a $x^2 - D'y^2$ si los impares que representa son congruentes con ± 1 módulo 8 y es equivalente a $5(x^2 - D'y^2)$ si los impares que representa son congruentes con ± 5 (mód 8).
- Si $D/4 \equiv 6$ (mód 8) llegamos a una conclusión similar.
- Si $D/4 \equiv 0$ (mód 8) entonces cada forma de (14.7) sólo representa a los impares congruentes con r módulo 8, luego determinan cuatro clases de formas diferentes.

Ejemplo Éstas son las formas cuadráticas reducidas de discriminante $D = -40$:

$$x^2 + 10y^2, \qquad 2x^2 + 5y^2.$$

Como $D' = -10 \equiv 6$ (mód 8), tenemos que la primera forma (que representa a 1) es equivalente módulo 2^n con $x^2 + 10y^2$ y con $-5(x^2 + 10y^2)$, mientras que la segunda (que representa a 5) es equivalente módulo 2^n con $-(x^2 + 10y^2)$ y con $5(x^2 + 10y^2)$. ∎

Así pues, para determinar las clases de equivalencia módulo 2^n (con $n \geq 3$) de formas cuadráticas de discriminante D par, sólo necesitamos estudiar los números impares que representan módulo 8. La diferencia con el caso p^n con n impar era que en ese caso sólo tenemos que distinguir si una forma dada f representa los restos cuadráticos o los restos no cuadráticos módulo p, para lo cual nos basta calcular el carácter $\chi_p(f)$, mientras que ahora tenemos cuatro clases en U_8 que una forma dada puede o no representar, para lo cual necesitamos considerar los caracteres módulo 8 definidos en la página 319. En términos del símbolo de Jacobi son

$$\delta(k) = \left(\frac{-1}{k}\right), \qquad \epsilon(k) = \left(\frac{2}{k}\right), \qquad (\delta\epsilon)(k) = \left(\frac{-2}{k}\right).$$

Notemos que, vistos como funciones en U_8, el carácter δ distingue a $\{1, 5\}$ de $\{-1, -5\}$, mientras que ϵ distingue a $\{1, -1\}$ de $\{5, -5\}$ y su producto $\epsilon\delta$ distingue a $\{1, -5\}$ de $\{-1, 5\}$.

Si f es una forma cuadrática de discriminante par D y a es cualquier número impar representado por f, definimos el *carácter* módulo 2 de f como

$$\chi_2(f) = \begin{cases} 1 & \text{si} \quad D/4 \equiv 1, 5 \text{ (mód 8)} \\ \epsilon(a) & \text{si} \quad D/4 \equiv 2 \text{ (mód 8)} \\ \delta(a) & \text{si} \quad D/4 \equiv 3, 4, 7 \text{ (mód 8)} \\ \delta(a)\epsilon(a) & \text{si} \quad D/4 \equiv 6 \text{ (mód 8)} \end{cases}$$

Si $D/4 \equiv 0$ (mód 8) definimos tres caracteres de f módulo 2, dados por

$$\chi_{21}(f) = \delta(a), \quad \chi_{22}(f) = \epsilon(a), \quad \chi_{23}(f) = \delta(a)\epsilon(a).$$

Recordemos además que para formas con discriminante impar hemos definido $\chi_2(f) = 1$.

Hemos demostrado que estos caracteres no dependen de la elección de a, así como que formas equivalentes módulo p^n (para $n \geq 3$) tienen el mismo carácter (o los mismos caracteres[3]) módulo p, para todo primo p, por lo que tiene sentido hablar del carácter de una clase de equivalencia de formas. Además tenemos el resultado siguiente:

Teorema 14.6 *Si p es primo, dos formas cuadráticas de discriminante D son equivalentes módulo p^n (para $n \geq 3$ si $p = 2$) si y sólo si tienen el mismo carácter módulo p. Esto ocurre siempre que $p \nmid D$.*

Notemos que si dos formas tienen el mismo carácter módulo 2 entonces son equivalentes módulo 2^n para todo n, incluyendo $n = 1, 2$, pues son equivalentes módulo 8, y esto implica que también lo son módulo 2 y 4. Lo que no es cierto es que dos formas equivalentes módulo 2 o 4 tengan necesariamente el mismo carácter y sean, pues, equivalentes módulo 2^n para todo n.

Para tratar unificadamente todos los casos en la medida de lo posible, conviene observar que para cada discriminante D y para cada primo p tenemos definido un carácter modular $\chi_p^* : U_p \longrightarrow \{\pm 1\}$ si p es impar, o $\chi_2^* : U_8 \longrightarrow \{\pm 1\}$ si $p = 2$, de manera que, para cada forma cuadrática f de discriminante D, se cumple que $\chi_p(f) = \chi_p^*\big([a]\big)$, donde a es cualquier número primo con p representado por f.

La función χ_p^* es constante igual a 1 si $p \nmid D$, es el símbolo de Legendre de p si $p \mid D$ es impar y es una de las funciones $1, \delta, \epsilon, \epsilon\delta$ si $p = 2$.

Combinando los resultados que hemos probado hasta ahora tenemos lo siguiente:

Teorema 14.7 *Si $n \geq 2$ es un número natural impar o bien $8 \mid n$ y f, g son dos formas cuadráticas de discriminante D, las afirmaciones siguientes son equivalentes:*

1. *f y g son equivalentes módulo n.*
2. *f y g representan los mismos enteros módulo n.*
3. *$\chi_p(f) = \chi_p(g)$ para todo primo $p \mid (n, D)$.*

[3]En lo sucesivo, cuando hablemos del carácter de una forma módulo un primo p habremos de recordar que si $p = 2$ puede haber en realidad tres caracteres, si bien no lo indicaremos explícitamente en cada ocasión para evitar constantes y monótonas salvedades como ésta.

DEMOSTRACIÓN: 1) ⇒ 2) es inmediato. Si suponemos 2), el teorema 13.26 nos da un entero a tal que $(a, D) = 1$ y que está representado por f. Por hipótesis g representa un entero $a' \equiv a \pmod{n}$.

Para cada primo $p \mid (n, D)$, se cumple $a' \equiv a \pmod{p}$ (y $a' \equiv a \pmod 8$ si $p = 2$), luego $\chi_p(f) = \chi_p^*(\bar{a}) = \chi_p^*(\bar{a}') = \chi_p(g)$, luego se cumple 3).

Si suponemos 3), entonces $\chi_p(f) = \chi_p(g)$ para todo $p \mid n$, pues esto se cumple trivialmente para los primos que no dividen a D, luego por 14.6 tenemos que f y g son equivalentes módulo p^e, donde e es el exponente de p en n, luego f y g son equivalentes módulo n por 14.2. ∎

Una última observación elemental es que si $\mathcal{O}$ es el anillo de enteros algebraicos de un cuerpo cuadrático y su discriminante D es par, entonces necesariamente $D/4 \equiv 2, 3 \pmod 4$, luego hay un único carácter χ_2 que es uno de los caracteres δ, ϵ o $\delta\epsilon$, pero no el carácter principal 1.

Ejemplos Consideremos de nuevo las cuatro clases de equivalencia de formas de discriminante 60 consideradas al inicio de este capítulo. Sus caracteres vienen dados por la tabla siguiente:

$D = 60$	χ_2	χ_3	χ_5
$x^2 - 15y^2$	+	+	+
$15x^2 - y^2$	−	−	+
$3x^2 - 5y^2$	−	+	−
$5x^2 - 3y^2$	+	−	−

Notemos que, como $D/4 = 15 \equiv 7 \pmod 8$, tenemos que $\chi_2 = \delta$. Por ejemplo, para $f = 3x^2 - 5y^2$, que representa al 3 y al -5, tenemos que

$$\chi_2(f) = \delta(3) = -1, \qquad \chi_3(f) = \left(\frac{-5}{3}\right) = \left(\frac{1}{3}\right) = 1, \qquad \chi_5(f) = \left(\frac{3}{5}\right) = -1.$$

El lector debería probar a calcular por sí mismo el resto de la tabla, así como las tablas de los ejemplos siguientes.

	$\delta\epsilon$	
$D = -40$	χ_2	χ_5
$x^2 + 10y^2$	+	+
$2x^2 + 5y^2$	−	−

	ϵ		
$D = -504$	χ_2	χ_3	χ_7
$x^2 + 126y^2$	+	+	+
$9x^2 + 14y^2$	+	−	+
$2x^2 + 63y^2$	+	−	+
$7x^2 + 18y^2$	+	+	+
$5x^2 + 4xy + 26y^2$	−	−	−
$10x^2 - 4xy + 13y^2$	−	+	−
$5x^2 - 4xy + 26y^2$	−	−	−
$10x^2 + 4xy + 13y^2$	−	+	−

	ϵ	
$D = -56$	χ_2	χ_7
$x^2 + 14y^2$	+	+
$2x^2 + 7y^2$	+	+
$3x^2 + 2xy + 5y^2$	−	−
$3x^2 - 2xy + 5y^2$	−	−

$D=-480$	χ_{21}	χ_{22}	χ_{23}	χ_3	χ_5
x^2+120y^2	+	+	+	+	+
$4x^2+4xy+31y^2$	−	+	−	+	+
$8x^2+15y^2$	−	+	−	−	−
$8x^2+8xy+17y^2$	+	+	+	−	−
$3x^2+40y^2$	−	−	+	+	−
$12x^2+12xy+13y^2$	+	−	−	+	−
$5x^2+24y^2$	+	−	−	−	+
$11x^2+2xy+11y^2$	−	−	+	−	+

$D=-120$	χ_2	χ_3	χ_5
x^2+30y^2	+	+	+
$2x^2+15y^2$	+	−	−
$3x^2+10y^2$	−	+	−
$5x^2+6y^2$	−	−	+

$D=480$	χ_{21}	χ_{22}	χ_{23}	χ_3	χ_5
$x^2+20xy-20y^2$	+	+	+	+	+
$4x^2+20xy-5y^2$	−	−	+	+	+
$3x^2+18xy-13y^2$	−	−	+	−	−
$-7x^2+12xy+12y^2$	+	+	+	−	−
$7x^2+12xy-12y^2$	−	+	−	+	−
$-3x^2+18xy+13y^2$	+	−	−	+	−
$-x^2+20xy+20y^2$	−	+	−	−	+
$-4x^2+20xy+5y^2$	+	−	−	−	+

$D=120$	χ_2	χ_3	χ_5
$x^2+10xy-5y^2$	+	+	+
$2x^2+8xy-7y^2$	+	−	−
$-2x^2+8xy+7y^2$	−	+	−
$-x^2+10xy+5y^2$	−	−	+

$D=-256$	χ_{21}	χ_{22}	χ_{23}
x^2+64y^2	+	+	+
$4x^2+4xy+17y^2$	+	+	+
$5x^2+2xy+13y^2$	+	−	−
$5x^2-2xy+13y^2$	+	−	−

Para el discriminante $D=-108$ tenemos tres clases de equivalencia de formas:

$$[x^2+27y^2],\quad [4x^2+2xy+7y^2],\quad [4x^2-2xy+7y^2]$$

y los dos caracteres χ_2 y χ_3 correspondientes a los divisores primos de D toman en ellas el valor 1 (en el caso de χ_2 por definición, pues $D/4\equiv 5$ (mód 8)).

Las tablas anteriores muestran muchos patrones que, por supuesto, no son casuales. ∎

14.2 Géneros

Definición 14.8 Diremos que dos formas cuadráticas de un mismo discriminante D son *del mismo género* si tienen los mismos caracteres.

Puesto que dos formas similares representan los mismos enteros y, por consiguiente, tienen los mismos caracteres, podemos hablar también de géneros de clases de equivalencia (estricta o no estricta) de formas cuadráticas de un discriminante dado.

Esto completa la clasificación de las formas cuadráticas binarias: éstas se dividen en órdenes según su discriminante, las formas de cada orden se dividen a su vez en géneros según sus caracteres, y las formas de un mismo género se distribuyen en clases de equivalencia (en $\mathbb{Z}$). Por último cada clase de equivalencia puede dividirse en dos clases de equivalencia estricta.

Por ejemplo, las tablas de caracteres que hemos mostrado en la sección precedente prueban que hay cuatro géneros de formas de discriminante 60, otros cuatro de discriminante -504, ocho géneros de discriminante -480 y otros ocho de discriminante 480.

Ejercicio: Comparar en cada caso el número de géneros que podría haber con el número de géneros que hay realmente. Por ejemplo, para $D = 480$ tenemos cinco caracteres, pero $\chi_{21}, \chi_{22}, \chi_{23}$ no son independientes (conociendo dos de ellos, conocemos el tercero), luego en principio tenemos cuatro caracteres que podrían tomar valores arbitrarios, lo que podría dar lugar hasta a 16 géneros, pero sólo hay 8.

Para ver una primera aplicación de la teoría de los géneros observamos el siguiente hecho trivial:

Teorema 14.9 *Si un entero n está representado por una forma cuadrática de discriminante D y $(n, D) = 1$, entonces n está representado concretamente por una forma de género G si y sólo si los valores $\chi_p^*(n)$ corresponden a los de G, para todo primo $p \mid D$.*

Como decimos, esto es trivial, pues si n está representado por la forma f, entonces $\chi_p(f) = \chi_p^*(n)$ para todo primo $p \mid D$, por lo que el género de f será el determinado por estos valores.

Ejemplo Consideremos las formas de discriminante $D = 120$. Podemos comprobar que los primos que se escinden en $\mathbb{Q}(\sqrt{30})$ son los que cumplen

$$p \equiv \pm 1, \pm 7, \pm 13, \pm 17, \pm 19, \pm 29, \pm 37, \pm 49 \text{ (mód 120)},$$

luego, según el teorema 12.29, un primo está representado por una de las cuatro formas cuadráticas[4]

$$x^2 - 30y^2, \quad 2x^2 - 15y^2, \quad 15x^2 - y^2, \quad 30x^2 - y^2$$

si y sólo si es $p = 2, 3, 5$ o cumple la condición anterior. En general, la dificultad está en determinar cuál o cuales de ellas representan a p, pero ahora la respuesta es inmediata: teniendo en cuenta que los caracteres χ_3 y χ_5 distinguen los cuatro géneros posibles, si un primo $p \neq 3, 5$ cumple estas condiciones, entonces

$$\begin{array}{lll}
p = x^2 - 30y^2 & \text{si y sólo si} & \chi_3^*(p) = 1, \quad \chi_5^*(p) = 1, \\
p = 2x^2 - 15y^2 & \text{si y sólo si} & \chi_3^*(p) = -1, \quad \chi_5^*(p) = -1, \\
p = 15x^2 - 2y^2 & \text{si y sólo si} & \chi_3^*(p) = 1, \quad \chi_5^*(p) = -1, \\
p = 30x^2 - y^2 & \text{si y sólo si} & \chi_3^*(p) = -1, \quad \chi_5^*(p) = 1.
\end{array}$$

Explícitamente:

$$\begin{array}{ll}
p = x^2 - 30y^2 & \text{si y sólo si } p \equiv 1, 4 \text{ (mód 15)}, \\
p = 2x^2 - 15y^2 & \text{si y sólo si } p \equiv 2, 8 \text{ (mód 15) o } p = 3, \\
p = 15x^2 - 2y^2 & \text{si y sólo si } p \equiv 7, 13 \text{ (mód 15)}, \\
p = 30x^2 - y^2 & \text{si y sólo si } p \equiv 11, 14 \text{ (mód 15) o } p = 5.
\end{array}$$

[4]Hemos tomado como representantes de las clases cuatro formas no reducidas, pero más sencillas.

o, si queremos incorporar la condición para que p sea representable por una forma de discriminante 120:

$$\begin{aligned}
&p = x^2 - 30y^2 \quad \text{si y sólo si } p \equiv -29, 1, 19, 49 \text{ (mód 120)},\\
&p = 2x^2 - 15y^2 \text{ si y sólo si } p \equiv -37, -13, -7, 17 \text{ (mód 120) o } p = 2, 3,\\
&p = 15x^2 - 2y^2 \text{ si y sólo si } p \equiv -17, 7, 13, 37 \text{ (mód 120)},\\
&p = 30x^2 - y^2 \quad \text{si y sólo si } p \equiv -49, -19, -1, 29 \text{ (mód 120) o } p = 5.
\end{aligned}$$

En este ejemplo ha sido fundamental que, las cuatro clases de equivalencia de formas de determinante -120 tienen géneros distintos, porque las condiciones que hemos dado determinan en realidad si un número dado está representado por una forma de un género determinado. Si hubiera varias clases de equivalencia de un mismo género (como sucede, por ejemplo, para $D = -504$ o $D = -108$), no sabríamos cuál o cuáles de ellas representan a p. ∎

Ejercicio: Caracterizar en términos de congruencias los primos que pueden expresarse en la forma $x^2 + 30y^2$, $2x^2 + 15y^2$, $3x^2 + 10y^2$ y $5x^2 + 6y^2$.

El razonamiento precedente no se aplica únicamente a primos, sino en realidad a cualquier entero p que cumpla $(p, 120) = 1$. Para eliminar esta restricción necesitamos relacionar los géneros con los grupos de clases.

Definición 14.10 Si M es un módulo cuyo anillo de coeficientes es un orden cuadrático $\mathcal{O}$ de discriminante D, para cada primo p definimos $\chi_p(M)$ como $\chi_p(f)$, donde f es cualquier forma cuadrática asociada a M por la correspondencia del teorema 12.19.

Obviamente $\chi_p(M)$ no depende de la forma elegida, porque dos formas asociadas a M son estrictamente equivalentes, luego son del mismo género. Más aún, dos módulos estrictamente similares tienen los mismos caracteres.

Como en el caso de las formas cuadráticas, diremos que dos módulos con el mismo anillo de coeficientes son *del mismo género* si tienen los mismos caracteres o, equivalentemente, si sus clases de equivalencia estricta de formas cuadráticas asociadas son del mismo género.

Obviamente, dos módulos estrictamente similares son del mismo género.

Ejemplo Consideremos el ideal $\mathfrak{a} = (\sqrt{3}) = \langle 3, \sqrt{3}\rangle$ del orden cuadrático $\mathcal{O} = \mathbb{Z}[\sqrt{3}]$. Su forma cuadrática asociada es

$$f(x, y) = \frac{\mathrm{N}(x3 + y\sqrt{3})}{\mathrm{N}(\mathfrak{a})} = \frac{9x^2 - 3y^2}{3} = 3x^2 - y^2,$$

que representa a -1, luego $\chi_3(\mathfrak{a}) = (-1/3) = -1$ y $\chi_2(\mathfrak{a}) = \delta(a) = -1$. Así pues, el género de $\mathfrak{a}$ es $(--)$, mientras que el género de $1 = (1) = \langle 1, \sqrt{3}\rangle$ es $(++)$. Los ideales 1 y $\mathfrak{a} = \sqrt{3}\cdot 1$ son similares, pero no son del mismo género, lo que prueba que no son estrictamente similares, y así tenemos un ejemplo de que dos módulos similares no son necesariamente del mismo género. ∎

El ejemplo anterior es un caso particular del teorema siguiente:

Teorema 14.11 *Sea $\mathcal{O}$ un orden cuadrático real y sea -1 la clase de similitud estricta formada por los módulos de la forma $\gamma\mathcal{O}$, donde γ es cualquier elemento de norma negativa. Entonces, para cada primo p, se cumple $\chi_p(-1) = \chi_p^*(-1)$.*

DEMOSTRACIÓN: Por el teorema 12.21, la clase de similitud estricta $1 = [\mathcal{O}]$ se corresponde con la clase de equivalencia estricta $[f]$ de la forma principal, y al estudiar la correspondencia entre módulos y formas vimos que si a una clase $[M]$ le corresponde la clase $[ax^2 + bxy + cy^2]$, a la clase $-[M]$ le corresponde $[-ax^2 + bxy - cy^2]$ luego, como la forma principal tiene $a = 1$, la clase -1 se corresponde con la clase de una forma con $a = -1$, luego representa -1, luego $\chi_p(-1) = \chi_p^*(-1)$. ∎

Como consecuencia, si un orden cuadrático $\mathcal{O}$ tiene discriminante $D > 0$ y existe un primo $p \mid D$ tal que $\chi_p^*(-1) = -1$, entonces en $\mathcal{O}$ las clases de similitud estricta 1 y -1 son distintas, pese a que sus módulos son similares, lo que implica que la unidad fundamental de $\mathcal{O}$ tiene que tener norma positiva.

En realidad podemos calcular el género de una clase de similitud estricta de módulos sin pasar por su clase de formas cuadráticas asociadas. Sólo necesitamos encontrar representantes que sean ideales de norma adecuada:

Teorema 14.12 *Sea $\mathcal{O}_m$ un orden cuadrático, sea $\mathfrak{a}$ un ideal de $\mathcal{O}_m$ de norma prima con m y sea p un primo que no divida a $\mathrm{N}(\mathfrak{a})$. Entonces $\chi_p(\mathfrak{a}) = \chi_p^*(\mathrm{N}(\mathfrak{a}))$.*

DEMOSTRACIÓN: Fijemos una base orientada de $\mathfrak{a} = \langle \alpha, \beta\rangle$, de modo que una forma cuadrática asociada a $\mathfrak{a}$ es

$$f(x,y) = \frac{\mathrm{N}(x\alpha + y\beta)}{\mathrm{N}(\mathfrak{a})}.$$

Por el teorema 13.11 (teniendo en cuenta que el anillo de coeficientes de $\mathfrak{a}$ es precisamente $\mathcal{O}_m$ por 13.24) tenemos que $\mathfrak{a} \mid \mathrm{N}(\mathfrak{a})$, luego $\mathrm{N}(\mathfrak{a}) \in \mathfrak{a}$. Por lo tanto existen enteros x_0, y_0 tales que $\mathrm{N}(\mathfrak{a}) = x_0\alpha + y_0\beta$, y por consiguiente $f(x_0, y_0) = \mathrm{N}(\mathrm{N}(\mathfrak{a}))/\mathrm{N}(\mathfrak{a}) = \mathrm{N}(\mathfrak{a})$. Concluimos que $\chi_p(\mathfrak{a}) = \chi_p(f) = \chi_p^*(\mathrm{N}(\mathfrak{a}))$. ∎

Ejemplo Consideremos el grupo de clases de similitud de formas de discriminante $D = -504$, cuya estructura determinamos en el ejemplo posterior al teorema 13.28:

$\mathcal{H}$			χ_2	χ_3	χ_7
1	$\langle 1, 3\sqrt{-14}\rangle$	$x^2 + 126y^2$	$+$	$+$	$+$
C	$\langle 5, 2 + 3\sqrt{-14}\rangle$	$5x^2 + 4xy + 26y^2$	$-$	$-$	$-$
C^2	$\langle 7, 3\sqrt{-14}\rangle$	$7x^2 + 18y^2$	$+$	$+$	$+$
C^3	$\langle 5, -2 + 3\sqrt{-14}\rangle$	$5x^2 - 4xy + 26y^2$	$-$	$-$	$-$
-1	$\langle 14, 3\sqrt{-14}\rangle$	$9x^2 + 14y^2$	$+$	$-$	$+$
$-C$	$\langle 10, -2 + 3\sqrt{-14}\rangle$	$10x^2 - 4xy + 13y^2$	$-$	$+$	$-$
$-C^2$	$\langle 2, 3\sqrt{-14}\rangle$	$2x^2 + 63y^2$	$+$	$-$	$+$
$-C^3$	$\langle 10, 2 + 3\sqrt{-14}\rangle$	$10x^2 + 4xy + 13y^2$	$-$	$+$	$-$

Aquí estamos llamando $C = [\langle 5, 2+3\sqrt{-14}\rangle]$ a la clase que en el ejemplo mencionado llamábamos $[\mathfrak{p}]$. Sobre este ejemplo podemos constatar que los caracteres son multiplicativos, es decir, que si C_1 y C_2 son dos clases de similitud de módulos, se cumple que $\chi_p(C_1C_2) = \chi_p(C_1)\chi_p(C_2)$. Esto es consecuencia inmediata del teorema anterior. ∎

Teorema 14.13 *Si M_1 y M_2 son módulos cuadráticos con un mismo anillo de coeficientes, para todo primo p, se cumple que $\chi_p(M_1M_2) = \chi_p(M_1)\chi_p(M_2)$.*

Demostración: Puesto que los caracteres dependen únicamente de las clases de similitud, no perdemos generalidad si suponemos que M_1 y M_2 son ideales de su anillo de coeficientes y que tienen norma prima con p (por el teorema 13.27). Entonces, por el teorema anterior,

$$\chi_p(M_1M_2) = \chi_p^*(\mathrm{N}(M_1M_2)) = \chi_p^*(\mathrm{N}(M_1)\,\mathrm{N}(M_2))$$

$$= \chi_p^*(\mathrm{N}(M_1))\chi_p^*(\mathrm{N}(M_2)) = \chi_p(M_1)\chi_p(M_2).$$ ∎

Ejercicio: Razonar que un carácter χ_p, o bien toma únicamente el valor $+1$, o bien toma el mismo número de veces el valor $+1$ que el valor -1.

Ejemplo Ahora ya podemos resolver el problema de cuándo un entero m no nulo esta representado por una forma concreta de discriminante $D = 120$. Por simplicidad fijaremos una de ellas, por ejemplo $2x^2-15y^2$, pero podemos razonar igualmente con las demás.

Empecemos con unas observaciones que son válidas, no sólo en este ejemplo en concreto, sino para todo discriminante D que tenga la propiedad de que no hay dos clases de similitud estricta de ideales que tengan el mismo género. Notemos que la clase principal contiene el ideal 1, cuyos caracteres son todos positivos, y es, por lo tanto, la única clase con esta propiedad. Por otro lado, si C es una clase arbitraria, como $\chi_p(C^2) = \chi_p(C)^2 = 1$, concluimos que todos los caracteres de C^2 son positivos, luego $C^2 = 1$. Equivalentemente $C = C^{-1}$.

Si un primo $p = \mathfrak{p}_1\mathfrak{p}_2$ se escinde, tomando clases vemos que $1 = [\mathfrak{p}_1][\mathfrak{p}_2]$, luego $[\mathfrak{p}_1] = [\mathfrak{p}_2]^{-1} = [\mathfrak{p}_2]$. Esto significa que todos los ideales de una misma norma prima pertenecen a la misma clase de similitud estricta.

Por lo tanto, si un ideal cumple $\mathrm{N}(\mathfrak{a}) = p^2$, con p primo, necesariamente, $[\mathfrak{a}] = 1$, pues, o bien p se conserva primo y $\mathfrak{a} = (p)$ es principal (generado por un elemento de norma positiva), en cuyo caso $[\mathfrak{a}] = 1$, o bien $\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2$, para dos ideales de norma p (iguales o distintos), pero entonces $[\mathfrak{a}] = [\mathfrak{p}_1]^2 = 1$.

Más en general, si $\mathrm{N}(\mathfrak{a}) = k^2r$, donde r es libre de cuadrados, podemos descomponer $\mathfrak{a} = a\mathfrak{b}\mathfrak{c}$, donde a es el producto de los divisores primos de $\mathfrak{a}$ correspondientes a primos racionales que se conservan, mientras que los divisores primos de $\mathfrak{a}$ de norma prima p los agrupamos en $\mathfrak{b}$ salvo si, para cada primo p, hay una cantidad impar de ellos, en cuyo caso uno lo dejamos en $\mathfrak{c}$. De este modo $\mathrm{N}(\mathfrak{b}) = b^2$ y los divisores primos $\mathfrak{b}$ se pueden agrupar en parejas de divisores de

la misma norma prima, por lo que la observación precedente nos da que $[\mathfrak{b}] = 1$, y trivialmente $[a] = 1$, ya que es un ideal principal generado por un elemento de norma positiva. Claramente $k^2 = a^2\,\mathrm{N}(\mathfrak{b})$ y $r = \mathrm{N}(\mathfrak{c})$, y además $[\mathfrak{a}] = [\mathfrak{c}]$.

Esto significa que una clase de similitud estricta C contiene un ideal de norma $m = k^2 r$ si y sólo si contiene un ideal de norma r.

A su vez, esto implica que una forma cuadrática f de discriminante D representa un entero no nulo $m = k^2 r$ si y sólo si[5] f representa a r.

En efecto, según el teorema 12.29 tenemos que f representa a m si y sólo si la clase de ideales C asociada a f cumple que C^{-1} contiene un ideal de norma m y, según acabamos de razonar, esto equivale a que C^{-1} (que es la propia C) contenga un ideal de norma r, es decir, que equivale a que f represente a r.

Pasamos ya a concretar la situación en el ejemplo que nos ocupa. Si llamamos

$$1 \leftrightarrow [x^2 - 30y^2], \quad C_1 \leftrightarrow [2x^2 - 15y^2], \quad C_2 \leftrightarrow [15x^2 - 2y^2], \quad C_3 \leftrightarrow [30x^2 - y^2]$$

a las clases de similitud asociadas a las clases de formas indicadas, es fácil ver que la operación del grupo de clases es

	1	C_1	C_2	C_3
1	1	C_1	C_2	C_3
C_1	C_1	1	C_3	C_2
C_2	C_2	C_3	1	C_1
C_3	C_3	C_2	C_1	1

(No es necesario comprobar nada. Sabemos que $C_1^2 = 1$, luego C_1C_2 no puede ser ni 1, ni C_1 y C_2, luego tiene que ser C_3, e igualmente en los demás casos.)

Queremos caracterizar los enteros representados por las formas de clase C_1. En primer lugar, que un entero $m = k^2 r$ esté representado por una forma cuadrática (adecuada) de discriminante 120 equivale a que en $\mathbb{Z}\big[\sqrt{30}\,\big]$ existan ideales de norma m, lo cual equivale a que los primos que dividan a r se ramifiquen o se escindan.

Por lo tanto, n está representado por una forma cuadrática de discriminante 120 si y sólo si los primos que dividen a r (los primos que dividen a n con exponente impar) son $p = 2, 3, 5$ o bien

$$p \equiv \pm1, \pm7, \pm13, \pm17, \pm19, \pm29, \pm37, \pm49 \text{ (mód 120)}.$$

Ahora tenemos que determinar cuál de las cuatro formas posibles representan a n si se da esta condición. Hemos razonado que será la misma forma que represente a r. Ahora necesitamos distinguir cuatro casos:

[5] Esto no es cierto en general, sino que depende de la hipótesis de que no hay dos clases de similitud del mismo género.

1. Si $3 \nmid r$, $5 \nmid r$, tenemos que n estará representado concretamente por $2x^2 - 15y^2$ si y sólo si $\chi_3^*(r) = \chi_5^*(r) = -1$.

2. Si $3 \mid r$, $5 \nmid r$, llamamos $r' = r/3$. Como la forma que representa a 3 es $2x^2 - 15y^2$, para que esta misma forma sea la que representa a $3r'$, es necesario y suficiente que las formas que representan a r' sean las asociadas a la clase 1, pues la clase 1 contiene un ideal de norma r' si y sólo si la clase C_1 contiene un ideal de norma $3r'$.

   Esto equivale a que $\chi_3^*(r') = \chi_5^*(r') = 1$.

3. Si $3 \nmid r$, $5 \mid r$, llamamos $r' = r/5$ y, por el mismo argumento, pero teniendo en cuenta que 5 está representado por las formas asociadas a C_3, concluimos que $2x^2 - 15y^2$ representa a r si y sólo si r' está representado por las formas asociadas a la clase C_2, ya que así C_2 contiene un ideal de norma r' si y sólo si $C_2C_3 = C_1$ contiene un ideal de norma r.

   Esto equivale a que $\chi_3^*(r') = 1$, $\chi_5^*(r') = -1$.

4. Si $15 \mid r$, llamamos $r' = r/15$ y, como C_1 contiene el ideal de norma 3 y C_3 contiene el ideal de norma 5, resulta que $C_2 = C_1C_3$ contiene el ideal de norma 15, luego para que $C_1 = C_2C_3$ contenga un ideal de norma r es necesario y suficiente que C_3 contenga un ideal de norma r', es decir que r' esté representado por las formas asociadas a la clase C_3.

   Esto equivale a que $\chi_r^*(r') = -1$, $\chi_5^*(r') = 1$.

En definitiva, la situación es la siguiente:

Un entero no nulo $m = k^2r$ (donde r es libre de cuadrados) es de la forma $m = 2x^2 - 15y^2$ los primos que dividen a r cumplen $p = 2, 3, 5$ o

$$p \equiv \pm1, \pm7, \pm13, \pm17, \pm19, \pm29, \pm37, \pm49 \text{ (mód 120)}.$$

y además se da uno de los casos siguientes:

1. $(r, 15) = 1$ *y* $r \equiv 2, 8$ (mód 15),
2. $3 \mid r$, $5 \nmid r$ *y* $r/3 \equiv 1, 4$ (mód 15),
3. $3 \nmid r$, $5 \mid r$ *y* $r/5 \equiv 7, 13$ (mód 15),
4. $15 \mid r$ *y* $r/15 \equiv 11, 14$ (mód 15).

Un planteamiento alternativo es determinar la clase de equivalencia que representa cada divisor primo de r según la conclusión del ejemplo de la página 510 y calcular la que representa al producto. Por ejemplo, para saber qué forma representa a $n = 4\,095 = 3^2 \cdot 5 \cdot 7 \cdot 13$, pasamos a $r = 5 \cdot 7 \cdot 13$, observamos que efectivamente todos los divisores primos son representables por formas de discriminante 120 y que, concretamente, están representados por las formas de las clases C_3, C_2 y C_2, respectivamente, luego el producto está representado por las formas de clase $C_3C_2^2 = C_3$, luego 4 095 no es de la forma $2x^2 - 15y^2$. ∎

Ejemplo 1 La teoría de los géneros no nos resuelve todos los problemas del tipo del que acabamos de analizar. Por ejemplo, si queremos saber qué números naturales son de la forma x^2+27y^2, nos encontramos con que hay tres clases de equivalencia estricta de formas cuadráticas de determinante -108, que admiten como representantes las formas

$$x^2+27y^2, \quad 4x^2+2xy+7y^2, \quad 4x^2-2xy+7y^2.$$

Las dos últimas son similares, por lo que representan los mismos enteros. El teorema 12.29 nos dice que una de estas formas representa un primo p si y sólo si el orden $\mathcal{O}_6$ del cuerpo $\mathbb{Q}\big(\sqrt{-3}\,\big)$ contiene ideales de norma p, lo que por 13.25, para $p \neq 2,3$ equivale a que $\mathcal{O}_1$ contenga ideales de norma p, es decir, a que p se escinda en $\mathcal{O}_1$, y esto sucede cuando $p \equiv 1 \pmod 3$ (véase el comentario tras el teorema 5.10).

Así pues, todo primo $p \equiv 1 \pmod 3$ es de la forma $p = x^2+27y^2$ o bien de la forma $p = 4x^2+2xy+7y^2$, pero ahora no podemos recurrir a la teoría de géneros para determinar cuáles son de cada tipo porque todas las formas cuadráticas de determinante -108 son del mismo género (véase el último ejemplo de la sección 14.1).

He aquí la lista de los primeros primos $p \equiv 1 \pmod 4$. Los destacados en negrita son los de la forma $p = x^2+27y^2$.

$$7,\ 13,\ 19,\ \mathbf{31},\ 37,\ \mathbf{43},\ 61,\ 67,\ 73,\ 79,\ 97,\ 103,\ \mathbf{109},$$

$$\mathbf{127},\ 139,\ 151,\ \mathbf{157},\ 163,\ 181,\ 193,\ 199,\ 211,\ \mathbf{223},\ \mathbf{229},\ \ldots$$

Puede probarse que no es posible distinguir unos de otros mediante congruencias, es decir, que no es cierto que los de una forma sean los congruentes con tales números módulo tal número y los de la otra forma los congruentes con tales otros números. Sin embargo, Euler conjeturó una caracterización muy simple. Nos ocuparemos de ella en el capítulo siguiente (véase la página 534). ∎

Ejemplo 2 Otro ejemplo similar al anterior lo proporcionan las formas cuadráticas de discriminante -256. Las tablas de la sección 14.1 muestran que hay cuatro clases de equivalencia estricta, pero que se dividen en dos géneros, con dos clases cada una. El género principal contiene a las formas

$$x^2+64y^2, \qquad 4x^2+4xy+17y^2,$$

mientras que el segundo género contiene dos clases de formas equivalentes, con representantes $5x^2 \pm 2xy + 13y^2$.

Según el teorema 12.29, un primo impar p está representado por una forma de discriminante -256 si y sólo si el orden $\mathcal{O}_{16}$ del anillo $\mathbb{Z}[i]$ contiene ideales de norma p, lo que por 13.25 equivale a que $\mathbb{Z}[i]$ contenga ideales de norma p. A su vez, esto equivale a que p se escinda en $\mathbb{Z}[i]$, es decir, a que $p \equiv 1 \pmod 4$.

La teoría de los géneros nos dice que para que p sea concretamente de la forma $p = 5x^2+2xy+13y^2$ es necesario y suficiente que además $\chi_{22}(p) = -1$, es decir que $p \equiv 3,5 \pmod 8$, lo que unido a la condición $p \equiv 1 \pmod 4$ equivale a que $p \equiv 5 \pmod 8$. Como $p = 2$ no es de esta forma, podemos concluir:

Un primo p es de la forma $p = 5x^2 + 2xy + 13y^2$ si y sólo si cumple que $p \equiv 5$ (mód 8).

Sin embargo, si $\chi_{22}(p) = 1$, es decir, si $p \equiv 1$ (mód 8), sólo sabemos que p es de la forma $p = x^2 + 64y^2$ o bien $p = 4x^2 + 4xy + 17y^2$, y no tenemos ningún criterio para saber cuál de los dos casos se da. La lista siguiente muestra los primeros casos. Los primos en negrita son los de la forma $x^2 + 64y^2$:

$$17,\quad 41,\quad \mathbf{73},\quad \mathbf{89},\quad 97,\quad \mathbf{113},\quad 137,\quad 193,\quad \mathbf{233},\quad 241,$$

$$\mathbf{257},\quad \mathbf{281},\quad 313,\quad \mathbf{337},\quad \mathbf{353},\quad 433,\quad 449,\quad 457.$$

Euler conjeturó una condición que los distingue (véase la página 559). ∎

Veamos otra aplicación de la teoría de géneros:

Teorema 14.14 *Sea $\mathcal{O}$ un orden cuadrático de discriminante $D < 0$ tal que cada género de ideales $\mathcal{O}$ consta de una única clase de similitud estricta y sea n un número natural tal que $(n, D) = 1$. Si n está representado por una forma cuadrática de discriminante D, el número total de tales representaciones es $u\sum_{r|n} \chi_k(r)$, donde χ_k es el carácter del cuerpo de cocientes de $\mathcal{O}$ y u es el número de unidades de $\mathcal{O}$.*

Demostración: Si $[f_1], \ldots, [f_h]$ son todas la clases de similitud estricta de formas cuadráticas de discriminante D, entonces n sólo está representado por una de ellas (digamos por $f = f_i$), pues si pudiera representarse por dos, como $(n, D) = 1$, podríamos calcular todos los caracteres de ambas formas usando n, y concluiríamos que ambas son del mismo género, en contra de lo supuesto.

Según el teorema 12.28, si la clase de equivalencia estricta $[f]$ se corresponde con la clase C de ideales de $\mathcal{O}$, las soluciones de la ecuación $f(x, y) = n$ se corresponden con los elementos de $\mathcal{O}$ de la forma $\xi = \xi_0\epsilon$, donde ϵ recorre las unidades de $\mathcal{O}$ (todas tienen norma 1 en un orden imaginario) y cada ξ posible se corresponde biunívocamente con un ideal de norma n en la clase C^{-1}.

Ahora bien, el número de unidades es finito (a partir del teorema 9.4 se sigue inmediatamente que las unidades de $\mathcal{O}$ son únicamente ± 1 salvo en los dos casos exceptuados en dicho teorema) y el número de ideales de norma n contenidos en la clase $C^{-}1$ coincide con el número de ideales de norma n de $\mathcal{O}$, pues si dos clases distintas contuvieran ideales de norma n, tendríamos dos formas no estrictamente equivalentes que representarían a n, y hemos visto que no es el caso.

Ahora basta tener en cuenta que, como n es primo con D, el teorema 13.25 nos da que el número de ideales de norma n de $\mathcal{O}$ es el mismo que el número de ideales de norma n del anillo de enteros algebraicos de k, que está calculado en el teorema 13.20, y ello nos lleva a la fórmula del enunciado. ∎

Nota Al final de la sección 4.3 calculamos el número de formas de expresar un número como suma de dos cuadrados. El resultado obtenido allí no coincide con el que proporciona el teorema anterior porque allí tuvimos el cuidado de

no distinguir expresiones que son "esencialmente" la misma. Por ejemplo, para $n = 5$ allí obtuvimos que $N_0 = 2$ y que el número de expresiones es $N_0/2 = 1$, mientras que el teorema anterior nos da $4(\chi_k(1) + \chi_k(5)) = 8$. La razón es que

$$5 = (\pm 1)^2 + (\pm 2)^2,$$

lo que nos da 8 expresiones distintas si contamos todas las combinaciones de signos y el orden de los sumandos, pero una sola si consideramos únicamente expresiones $x^2 + y^2$ con x, y no negativos y sin tener en cuenta el orden. ∎

De aquí se deduce un criterio de primalidad:

Teorema 14.15 *Sea $f(x, y)$ una forma cuadrática asociada a un orden de discriminante $D < -4$ en el que cada género contenga una única clase de similitud estricta de ideales. Sea p un número natural primo con D que se expresa exactamente de cuatro formas distintas como $p = f(x, y)$ en las que $(x, y) = 1$. Entonces p es primo.*

DEMOSTRACIÓN: El orden de f tendrá exactamente dos unidades, luego el teorema anterior junto con 13.20 nos da que su cuerpo cuadrático tiene exactamente dos ideales de norma p. Más aún, la demostración de 13.20 muestra que esto sucede porque p es divisible entre un único divisor primo que se escinde y además con multiplicidad 1. Basta ver que p no es divisible entre primos que se conservan o se ramifican. Ciertamente, p no es divisible entre primos que se ramifican, pues por hipótesis es primo con el discriminante del cuerpo. Supongamos que q es un primo que se conserva y divide a p.

Consideremos un módulo asociado a la forma f. Podemos exigir que sea un ideal de norma prima con q. Más aún, según el teorema 13.8 podemos tomarlo de la forma $\mathfrak{a} = \langle a, b + m\omega\rangle$, donde $\mathrm{N}(\mathfrak{a}) = a$. Cambiando f por una forma estrictamente equivalente, podemos suponer que

$$p = f(x, y) = \frac{\mathrm{N}(ax + (b + m\omega)y)}{a}.$$

Notemos que si $(x, y) = 1$ y aplicamos un cambio de variables lineal de determinante 1, las imágenes x', y' siguen cumpliendo lo mismo. (Si x' e y' fueran divisibles entre un mismo primo, aplicando el cambio de variables inverso obtendríamos que lo mismo vale para x, y.) El numerador es un entero racional, luego tenemos que $q \mid \mathrm{N}(ax + (b + m\omega)y)$, y como q es primo en el orden cuadrático, también $q \mid ax + (b + m\omega)y$. Esto implica que $q \mid ax + by$, $q \mid my$, con lo que $q \mid y$ y $q \mid ax$, luego $q \mid x$, lo cual es imposible. ∎

Números idóneos Euler conocía un caso particular de este teorema, que podemos expresar en términos de lo que llamó *números idóneos*. Euler los definió en otros términos, pero en nuestro contexto podemos definirlos como los números n tales que cada género de discriminante $-4n$ contiene una única clase de similitud.

Tabla 14.1: Los números idóneos de Euler, agrupados por el número de clases del orden de discriminante $-4n$

h	
1	$1, 2, 3, 4, 7$
2	$5, 6, 8, 9, 10, 12, 13, 15, 16, 18, 22, 25, 28, 37, 58$
4	$21, 24, 30, 33, 40, 42, 45, 48, 57, 60, 70, 72, 78, 85, 88, 93, 102, 112,$ $130, 133, 177, 190, 232, 253$
8	$105, 120, 165, 168, 210, 240, 273, 280, 312, 330, 345, 357, 385, 408,$ $462, 520, 760$
16	$840, 1\,320, 1\,365, 1\,848$

Euler encontró los 65 números idóneos que muestra la tabla 14.1. Se conjetura que no hay más, y se ha demostrado que a lo sumo puede haber uno más, y que esto no sucede si se cumple una generalización de la hipótesis de Riemann. Vemos que los menores números que no son idóneos son $n = 11, 14, 17, 19$.

Si particularizamos el teorema anterior a los números idóneos resulta el criterio siguiente:

Teorema 14.16 *Si n es un número idóneo y p es un número impar que se expresa de forma única como $p = x^2 + ny^2$, para ciertos números naturales x, y tales que $(x, ny) = 1$, entonces p es primo.*

Las cuatro representaciones de las que habla este teorema son entonces $(\pm x, \pm y)$. Euler usó este criterio para encontrar primos grandes. El mayor que encontró fue

$$18\,518\,809 = 197^2 + 1\,848 \cdot 100^2,$$

(nosotros hemos comprobado la unicidad en la página 156, donde hemos dejado como ejercicio comprobar que el criterio también es aplicable al número 142 969.) aunque sería más acertado decir que fue el mayor que se molestó en comprobar, pues lo cierto es que todos los primos que se escinden en los órdenes cuadráticos de discriminante $-4n$, donde n es un número idóneo, pueden obtenerse de este modo, por lo que estamos hablando de infinitos primos.

Por ejemplo, con un esfuerzo asumible, Euler podría haber demostrado[6] que

$$184\,899\,940\,009 = 9\,997^2 + 1\,948 \cdot 10\,000^2$$

es primo, pues, usando tablas de índices como se muestra en el ejemplo de la página 156, sólo es necesario realizar una criba con congruencias desde $p = 5$ hasta $p = 61$ (14 primos en total). Una comprobación directa requeriría probar a dividir el número entre los primeros 36 162 primos, hasta el 429 991. ∎

[6]Puede objetarse que antes es necesario elegir con qué número probar y tener la suerte de que realmente sea primo, pero si fijamos $y = 10\,000$ y buscamos valores de $x < 10\,000$, tenemos que descartar $x = 9\,999$, porque el número resultante sería múltiplo de 3 y $x = 9\,998$ porque sería par, así que $x = 9\,997$ es el primer candidato que podría dar lugar a un primo, y resulta que lo es. Más aún, si seguimos descendiendo, se descartan obviamente $x = 9\,996$, $9\,995$, $9\,994$, $9\,993$, $9\,992$ y el siguiente valor, $x = 9\,991$ también da lugar a un primo.

14.3 El número de géneros

Las tablas de caracteres que hemos calculado muestran que en cada caso sólo se dan realmente la mitad de las combinaciones de caracteres posibles. Por ejemplo, en el caso $D = 60$ se da la combinación $(-,-,+)$, pero no la combinación $(+,-,+)$, y de las 8 combinaciones que podrían darse en principio, sólo se dan 4. Este fenómeno se debe esencialmente al teorema siguiente, que es, de hecho, equivalente a la ley de reciprocidad cuadrática:

Teorema 14.17 *Si k es un cuerpo cuadrático y M es un módulo cuyo anillo de coeficientes es el anillo de enteros $\mathcal{O}$ de k, entonces*

$$\prod_p \chi_p(M) = 1,$$

es decir, el número de caracteres negativos de M es par.

Demostración: Vamos a probar un hecho general que volveremos a usar más adelante, y es que si q es un primo impar que no divide al discriminante Δ de k, entonces

$$\prod_p \chi_p^*(q) = \left(\frac{\Delta}{q}\right) = \chi_k(q),$$

donde χ_k es el carácter de k (definición 9.14). En efecto, sea $k = \mathbb{Q}\big(\sqrt{d}\,\big)$ y distingamos varios casos:

1) Si $d \equiv 1 \pmod 4$, entonces $\Delta = d$ y todos los primos $p \mid \Delta$ son impares. Por lo tanto,

$$\prod_p \chi_p^*(q) = \prod_p \left(\frac{q}{p}\right) = \left(\frac{q}{|\Delta|}\right) = \begin{cases} \left(\frac{\Delta}{q}\right) & \text{si } \Delta > 0, \\ \left(\frac{-1}{q}\right)\left(\frac{|\Delta|}{q}\right) = \left(\frac{\Delta}{q}\right) & \text{si } \Delta < 0. \end{cases}$$

2) Si $d \equiv -1 \pmod 4$, entonces $\Delta = 4d$ y $\chi_2^*(q) = \delta(q) = (-1/q)$. Por lo tanto,

$$\prod_p \chi_p^*(q) = \left(\frac{q}{|\Delta|/4}\right)\left(\frac{-1}{q}\right) = \begin{cases} \left(\frac{\Delta/4}{q}\right) = \left(\frac{\Delta}{q}\right) & \text{si } \Delta > 0, \\ \left(\frac{-\Delta/4}{q}\right)\left(\frac{-1}{q}\right) = \left(\frac{\Delta}{q}\right) & \text{si } \Delta < 0. \end{cases}$$

3) Si $d = 2d'$, entonces $\Delta = 8d'$ y caben dos posibilidades:

3a) Si $d' \equiv 1 \pmod 4$, entonces $d \equiv 2 \pmod 8$, luego $\chi_2^*(q) = \epsilon(q) = (2/q)$. Entonces

$$\prod_p \chi_p^*(q) = \left(\frac{q}{|d'|}\right)\left(\frac{2}{q}\right) = \begin{cases} \left(\frac{d'}{q}\right)\left(\frac{2}{q}\right) = \left(\frac{\Delta}{q}\right) & \text{si } \Delta > 0, \\ \left(\frac{-d'}{q}\right)\left(\frac{-1}{q}\right)\left(\frac{2}{q}\right) = \left(\frac{\Delta}{q}\right) & \text{si } \Delta < 0. \end{cases}$$

3b) Si $d' \equiv 3 \pmod 4$, entonces $d \equiv 6 \pmod 8$, $\chi_2^*(q) = \delta(q)\epsilon(q) = (-2/q)$.

$$\prod_p \chi_p^*(q) = \left(\frac{q}{|d'|}\right)\left(\frac{-2}{q}\right) = \begin{cases} \left(\frac{-d'}{q}\right)\left(\frac{-1}{q}\right)\left(\frac{-2}{q}\right) = \left(\frac{\Delta}{q}\right) & \text{si } \Delta > 0, \\ \left(\frac{-d'}{q}\right)\left(\frac{-2}{q}\right) = \left(\frac{\Delta}{q}\right) & \text{si } \Delta < 0. \end{cases}$$

Pasamos ya a demostrar el teorema. Por 13.27 sabemos que M es estrictamente similar a un ideal de $\mathcal{O}$ de norma prima con el discriminante Δ de k. Puesto que los módulos estrictamente similares tienen los mismos caracteres, no perdemos generalidad si suponemos que M es un ideal en estas condiciones. Más aún, dicho ideal se descompone en producto de ideales primos $\mathfrak{q}$ de norma impar prima con Δ, luego basta probar el teorema para cada ideal $\mathfrak{q}$. Sea q el primo racional que cumple $\mathfrak{q} \mid q$, de modo que $q \nmid \Delta$.

Si $p \mid \Delta$, el teorema 14.12 nos da que $\chi_p(\mathfrak{q}) = \chi_p^*(\mathrm{N}(\mathfrak{q}))$. Si $\mathrm{N}(\mathfrak{q}) = q^2$ trivialmente se cumple que $\chi_p(\mathfrak{q}) = 1$ para todo p, luego el producto también vale 1. Suponemos, pues, que $\mathrm{N}(\mathfrak{q}) = 1$, lo que equivale a que el primo q se escinde en $\mathcal{O}$ (ya que para ramificarse debería dividir a Δ). Entonces, según lo que hemos demostrado:

$$\prod_p \chi_p(\mathfrak{q}) = \prod_p \chi_p^*(q) = \left(\frac{\Delta}{q}\right) = \chi_k(q) = 1.$$

■

Gauss demostró una versión equivalente del teorema anterior (en términos de formas cuadráticas) sin usar la ley de reciprocidad cuadrática y a partir de ella obtuvo otra prueba de la ley de reciprocidad.

Esto prueba que en los grupos de clases de equivalencia estricta de formas cuadráticas cuyo discriminante coincida con el de un cuerpo cuadrático, el número de géneros es a lo sumo la mitad del número que podrían definir los caracteres. Luego veremos que es exactamente la mitad, pero antes vamos a ver lo que sucede en los grupos correspondientes a órdenes cuadráticos no maximales. Si miramos tablas como las correspondientes a $D = -504$ o ± 480, vemos que ya no es cierto que el número de caracteres negativos de un género deba ser par. Sin embargo, vamos a ver que el teorema anterior también tiene su repercusión en este caso.

Caracteres fundamentales Consideremos el grupo $\mathcal{H}_m$ de clases de similitud estricta de módulos cuyo anillo de coeficientes es el orden $\mathcal{O}_m$ de un cuerpo cuadrático k de discriminante Δ_k, de modo que el discriminante de $\mathcal{O}_m$ es $D = m^2\Delta_k$. Sea $\mathcal{H}_1$ el grupo de clases de similitud estricta de módulos con anillo de coeficientes $\mathcal{O}_1$.

El teorema 13.28 define una aplicación $\mathcal{H}_m \longrightarrow \mathcal{H}_1$ dada por $[\mathfrak{a}] \mapsto [\mathfrak{a}^*]$, donde $\mathfrak{a} \mapsto \mathfrak{a}^*$ es la aplicación dada por el teorema 13.25. Recordemos que conserva la norma. Esto hace que si p es un primo impar que divide a Δ_k, entonces

$$\chi_p([\mathfrak{a}]) = \chi^*(\mathrm{N}(\mathfrak{a})) = \left(\frac{\mathrm{N}(\mathfrak{a})}{p}\right) = \left(\frac{\mathrm{N}(\mathfrak{a}^*)}{p}\right) = \chi_p([\mathfrak{a}^*]).$$

Lo mismo vale para $p = 2$ (suponiendo que $p \mid \Delta_k$) si $D/4 \equiv 2, 3, 4, 6, 7$ (mód 8), pues esto sucede cuando m es impar, y entonces $m^2 \equiv 1$ (mód 8), por lo que $D/4 \equiv \Delta_k/4$ (mód 8) y χ_2^* es el mismo para $\mathcal{O}_m$ y para $\mathcal{O}_1$.

Definición 14.18 Llamaremos *caracteres fundamentales* de un orden $\mathcal{O}_m$ de un cuerpo cuadrático k como los caracteres χ_p del grupo $\mathcal{H}_m$ de clases de similitud estricta de módulos de $\mathcal{O}_m$ correspondientes a los primos $p \mid \Delta_k$, salvo en el caso en que el discriminante D de $\mathcal{O}_m$ cumpla $D/4 \equiv 0$ (mód 8), en el cual hay tres caracteres $\chi_{21}, \chi_{22}, \chi_{23}$ y sólo consideraremos fundamental al que cumple que χ_{2i}^* coincide con el carácter χ_2^* de $\mathcal{O}_1$.

El razonamiento precedente (y el convenio que acabamos de adoptar en el último caso) justifican que los caracteres fundamentales toman los mismos valores sobre cada clase del grupo $\mathcal{H}_m$ que sobre su imagen en $\mathcal{H}_1$, de modo que las combinaciones de caracteres fundamentales que se dan realmente en $\mathcal{H}_m$ coinciden con las combinaciones de caracteres que se dan en $\mathcal{H}_1$.

En particular, el teorema 14.17 nos da ahora que, en cualquier grupo de clases de similitud estricta de módulos, cada clase tiene un número par de caracteres fundamentales negativos.

Por ejemplo, ahora podemos precisar que, en la tabla de caracteres correspondiente al discriminante $D = -504$, los caracteres fundamentales son χ_2 y χ_7, y en efecto, sólo toman los valores $(+,+)$ y $(-,-)$. En el caso del discriminante $D = 480$, el carácter χ_2^* de discriminante $\Delta_k = 120$ es $\delta\epsilon$, luego los caracteres fundamentales son χ_{23}, χ_3, χ_5, como podríamos haber deducido simplemente de mirar la tabla, pues χ_{23} es el único de los tres caracteres asociados a 2 que cumple que, unido a χ_3 y χ_5 el número de caracteres negativos de cualquier clase es par.

Con esto tenemos probado que, en cualquier grupo de clases de similitud estricta, las combinaciones de caracteres que se dan realmente son a lo sumo la mitad de las que podrían definir los caracteres (o la cuarta parte cuando hay tres caracteres módulo 2, ya que obviamente $\chi_{21}\chi_{22}\chi_{23} = 1$). Las tablas de caracteres sugieren que no hay más restricciones y, en efecto, vamos a demostrar que es así:

Teorema 14.19 *Sea $\mathcal{O}$ un orden cuadrático. Entonces una combinación de caracteres (no triviales) se corresponde con un género de $\mathcal{O}$ si y sólo si el número de caracteres fundamentales negativos es par y, en caso de que haya tres caracteres módulo 2, el número de caracteres negativos módulo 2 es par.*

Demostración: Sea k el cuerpo cuadrático al que pertenece $\mathcal{O}$. Puesto que los valores de $\chi_p^*(x)$ dependen sólo del resto de x módulo p (o módulo 8), el teorema chino del resto nos da un entero impar m primo con el discriminante D de $\mathcal{O}$ tal que $\chi_p^*(m)$ toma cualquier juego de valores prefijado, y m está determinado módulo D (aquí se usa la restricción sobre los caracteres módulo 2). Si probamos que $\mathcal{O}$ tiene un ideal de norma m, evidentemente su género tendrá la combinación de caracteres prefijada.

No es fácil probar la existencia de tal ideal, así que simplificaremos el problema haciendo uso del teorema de Dirichlet sobre primos en progresiones aritméticas. La sucesión $m+s\Delta$ contiene un primo q, de modo que podemos razonar con q en lugar de m. Ahora basta observar que en la demostración de 14.17 hemos visto que

$$\chi_k(q) = \prod_{p|\Delta_k} \chi_p^*(q) = 1,$$

donde usamos la hipótesis sobre el número par de caracteres fundamentales negativos, y esto significa que q se escinde en k, luego existe un primo $\mathfrak{q}$ de norma q, y como $(q, D) = 1$, la correspondencia entre los ideales de $\mathcal{O}$ y los del orden maximal de k implica que $\mathcal{O}$ también tiene un primo de norma q. ∎

Así pues si un orden cuadrático $\mathcal{O}$ tiene r caracteres, entonces el número de géneros de dicho orden es 2^{r-1} o bien 2^{r-2} si hay tres caracteres módulo 2.

Capítulo XV

La ley de reciprocidad cúbica

Igual que hemos estudiado los restos cuadráticos módulo primos, podemos estudiar los restos cúbicos. Euler ya hizo algunas conjeturas sobre cuándo 2 y otros enteros pequeños son o no restos cúbicos módulo un primo, y Gauss llegó a conjeturar una ley de reciprocidad cúbica análoga a la ley de reciprocidad cuadrática, pero la primera demostración publicada se debe a Eisenstein, y es de 1844. Jacobi enunció varios resultados sobre restos cúbicos, pero no publicó ninguna prueba (si bien acusó a Eisenstein de haberle plagiado).

15.1 Restos cúbicos

Los resultados básicos sobre los restos cúbicos módulo un primo $p \neq 3$ se deducen del teorema 3.37. Vamos a enunciarlo en este caso particular:

Teorema 15.1 *Si un primo p cumple $p \equiv 1 \pmod 3$, entonces sólo la tercera parte de las clases de U_p son cubos, mientras que si $p \equiv -1 \pmod 3$ o $p = 3$, entonces todas las clases de U_p son cubos.*

Demostración: Basta aplicar 3.37 teniendo en cuenta que $d = (3, p-1)$ es igual a 1 si $p = 3$ o $p \equiv -1 \pmod 3$ y es igual a 3 cuando $p \equiv 1 \pmod 3$. ∎

Así pues, el problema de distinguir los restos cúbicos y los restos no cúbicos módulo primos $p \equiv -1 \pmod 3$ es trivial, ya que todos los enteros son cubos módulo p (trivialmente en el caso de los múltiplos de p), luego el problema sólo tiene interés para los primos $p \equiv 1 \pmod 3$.

Euler llegó a formular algunas conjeturas en este terreno. Si comparamos con el caso cuadrático, vemos que 2 es un resto cuadrático módulo un primo p si y sólo si $p \equiv \pm 1 \pmod 8$, o que 3 es un resto cuadrático módulo p si y sólo si $p \equiv \pm 1 \pmod{12}$ (teorema 5.10), o que 5 es un resto cuadrático módulo p si y sólo si $p \equiv \pm 1 \pmod 5$, etc. Sin embargo, no existe ninguna condición de este estilo que determine, por ejemplo, el carácter cúbico de 2 módulo un primo $p \equiv 1 \pmod 3$. El lector escéptico puede tratar de encontrar alguna a partir de

la tabla siguiente, que contiene los primeros primos congruentes con 1 módulo 3, de modo que 2 es un resto cúbico módulo los destacados en negrita:

$$7,\ 13,\ 19,\ \mathbf{31},\ 37,\ \mathbf{43},\ 61,\ 67,\ 73,\ 79,\ 97,\ 103,\ \mathbf{109},$$

$$\mathbf{127},\ 139,\ 151,\ \mathbf{157},\ 163,\ 181,\ 193,\ 199,\ 211,\ \mathbf{223},\ \mathbf{229},\ \ldots$$

Si tiene buena memoria, tal vez recuerde haber visto antes esta misma lista. Los primos marcados en negrita coinciden con los primos de la forma $p = x^2 + 27y^2$. Esto fue lo que conjeturó Euler, que un primo $p \equiv 1$ (mód 3) es de esta forma si y sólo si 2 es un resto cúbico módulo p. Veremos que esto es consecuencia de la ley de reciprocidad cúbica.

Para enunciarla definiremos un análogo cúbico del símbolo de Legendre, y en este punto conviene observar que sería un error definirlo como

$$\left(\frac{n}{p}\right)_3 = \begin{cases} 1 & \text{si } n \text{ es un resto cúbico módulo } p, \\ -1 & \text{en caso contrario.} \end{cases}$$

Consideremos el caso concreto $p = 13$. Los cubos módulo 13 son:

$$p = 13$$

x	1	2	3	4	5	6	7	8	9	10	11	12
x^3	1	8	1	12	8	8	5	5	1	12	5	12

por lo que, de las 12 clases de restos módulo 13, sólo son cubos $1, 5, 8, 12$. Con la definición anterior, tendríamos que

$$\left(\frac{2}{13}\right)_3 \left(\frac{3}{13}\right)_3 = (-1)(-1) = 1 \neq -1 = \left(\frac{2\cdot 3}{13}\right)_3.$$

Así pues, el "símbolo cúbico" así definido no sería multiplicativo, y esto lo volvería completamente inútil.

Gauss comprendió que no es razonable dividir las 12 clases de restos módulo 13 en dos categorías, una con los tres restos cúbicos y otra con los seis restos no cúbicos, y mucho menos tratar de distinguirlas mediante un signo positivo o negativo, porque en realidad hay tres categorías de restos, una de las cuales es la formada por los tres restos cúbicos, pero luego hay otras dos, cada una de ellas con tres restos no cúbicos. Y a la hora de distinguirlas, el 1 y el -1 (las raíces cuadradas de la unidad que usamos para distinguir los restos cuadráticos de los restos no cuadráticos) deben ser sustituidos por las raíces cúbicas de la unidad, $1, \zeta$ y ζ^2, de modo que el análogo "correcto" del símbolo de Legendre debe tomar estos tres valores.

Enteros de Eisenstein Antes de profundizar en estas ideas vamos a recordar los hechos básicos sobre los enteros de Eisenstein, que van a representar un papel esencial en la prueba de la ley de reciprocidad cúbica.

Sabemos que $\mathbb{Z}[\zeta]$ es un dominio euclídeo (en particular, un dominio de ideales principales y un dominio de factorización única) con seis unidades:

$$\pm 1, \qquad \pm\zeta, \qquad \pm\zeta^2,$$

que el único primo racional que se ramifica es $3 = -\zeta^2\lambda^2$, donde $\lambda = 1 - \zeta$, y los demás se dividen entre los que se escinden (los que cumplen $p \equiv 1$ (mód 3)) y los que se conservan (los que cumplen $p \equiv -1$ (mód 3)). Recordemos también que la norma de un entero de Eisenstein viene dada por

$$\mathrm{N}(a + b\zeta) = a^2 + b^2 - ab.$$

Si α es un entero de Eisenstein, el teorema 13.9 (teniendo en cuenta la observación precedente, según la cual $\mathrm{N}((\alpha)) = |\,\mathrm{N}(\alpha)|$), sabemos que hay exactamente $\mathrm{N}(\alpha)$ clases de restos módulo α.

En particular, hay 9 clases de restos módulo 3, que son las siguientes:

0	1	2
ζ	$1+\zeta$	$2+\zeta$
2ζ	$1+2\zeta$	$2+2\zeta$

o equivalentemente, teniendo en cuenta que $1 + \zeta + \zeta^2 = 0$,

0	1	-1
ζ	$-\zeta^2$	$-\lambda$
$-\zeta$	λ	ζ^2

Vemos así, que el grupo de las unidades del anillo $\mathbb{Z}[\zeta]/(3)$ consta de 6 clases, que admiten como representantes a las seis unidades de $\mathbb{Z}[\zeta]$. Las otras tres clases son 0 y $\pm\bar{\lambda}$.

Concluimos que todo entero de Eisenstein α no divisible entre λ es congruente módulo 3 con una única unidad de $\mathbb{Z}[\zeta]$, y multiplicando por la unidad inversa concluimos que todo entero de Eisenstein no divisible entre λ tiene un único asociado congruente con 1 módulo 3. Esto justifica la definición siguiente:

Definición 15.2 Un entero de Eisenstein α no divisible entre λ es *primario* si $\alpha \equiv \pm 1$ (mód 3).

Acabamos de probar que todo entero de Eisenstein α no divisible entre λ tiene exactamente dos asociados primarios (uno congruente con 1 y otro con -1 módulo 3).

En la definición de entero primario podríamos haber exigido que α fuera congruente necesariamente con 1 (o con -1), y así cada entero primo con λ tendría un único asociado primario,[1] pero para lo que vamos a hacer resulta más práctico este convenio más laxo.

[1] Por razones técnicas, el convenio usual cuando se adopta una definición estricta es exigir que los enteros primarios sean congruentes con -1 módulo 3.

Distinguir asociados primarios es el equivalente para los enteros de Eisenstein a lo que hacemos inadvertidamente al tratar con enteros ordinarios al tomarlos positivos. Por ejemplo, el enunciado usual de la ley de reciprocidad cuadrática para un par de primos p y q requiere que éstos sean positivos. Veremos que el enunciado correspondiente a la ley de reciprocidad cúbica requerirá que los primos considerados sean primarios.

Ejemplo Consideremos $\pi = \zeta - 2$, que es un primo de norma

$$\mathrm{N}(\pi) = 1^2 + (-2)^2 - 1(-2) = 7.$$

Según las tablas anteriores, vemos que

$$\pi \equiv 1 + \zeta \equiv -\zeta^2 \text{ (mód 3)},$$

luego sus asociados primarios se obtienen multiplicando por $\pm\zeta$, y son

$$\zeta^2 - 2\zeta = -3\zeta - 1, \qquad 3\zeta + 1.$$

En general, sus seis asociados son:

$$\zeta - 2, \quad -\zeta + 2, \quad \zeta^2 - 2\zeta = -3\zeta - 1, \quad 3\zeta + 1, \quad \zeta^3 - 2\zeta^2 = 2\zeta + 3, \quad -2\zeta - 3.$$

■

Más adelante necesitaremos el siguiente hecho elemental:

Teorema 15.3 *Un entero de Eisenstein no nulo ni unitario es primario si y sólo si se descompone en producto de factores primos primarios.*

Demostración: Es inmediato que el producto de enteros primarios es primario, luego una implicación es obvia. Si α es primario, podemos descomponerlo como $\alpha = \epsilon\pi_1 \cdots \pi_r$, donde los π_i son primos primarios y ϵ es una unidad. Puesto que tanto α como los π_i son congruentes con ± 1 mód 3, podemos concluir que $\epsilon \equiv \pm 1$ (mód 3), pero las únicas unidades que cumplen esto son $\epsilon = \pm 1$, luego cambiando π_1 por $\epsilon\pi_1$ tenemos una descomposición de α en factores primos primarios. ■

Observemos ahora que, según vimos en la sección 6.2, podemos expresar

$$\zeta = \frac{-1 + \sqrt{-3}}{2},$$

con lo que todo entero de Eisenstein $\alpha \equiv a + b\zeta$ puede expresarse en la forma

$$\alpha = a + b\frac{-1 + \sqrt{-3}}{2} = \frac{2a - b + b\sqrt{-3}}{2}.$$

La congruencia $\alpha \equiv \pm 1$ (mód 3) equivale a que $a \equiv \pm 1$ (mód 3) y $b \equiv 0$ (mód 3), y en tal caso

$$\alpha = \frac{L + 3M\sqrt{-3}}{2},$$

donde $L = 2a - b$, $M = b/3$. Recíprocamente, dado un número α de esta forma, su expresión en la forma $\alpha = a + b\zeta$ viene dada por

$$2a = L + 3M, \qquad b = 3M.$$

Será entero si $2 \mid L + 3M$ (lo que equivale a que $L \equiv M$ (mód 2)) y cumplirá la condición $\alpha \equiv \pm 1$ (mód 3) si además $L \equiv \mp 1$ (mód 3). Además

$$\mathrm{N}(\alpha) = \frac{L^2 + 27M^2}{4}.$$

En particular, si $p \equiv 1$ (mód 3) es un primo racional, sabemos que se escinde en $\mathbb{Z}[\zeta]$, es decir, que es de la forma $p = \pi\bar{\pi}$, donde los factores primos están unívocamente determinados si exigimos que $\pi \equiv \bar{\pi} \equiv 1$ (mód 3) o, alternativamente, que $\pi \equiv \bar{\pi} \equiv -1$ (mód 3). Entonces

$$\pi = \frac{L + 3M\sqrt{-3}}{2}, \qquad \bar{\pi} = \frac{L - 3M\sqrt{-3}}{2}\zeta \qquad p = \frac{L^2 + 27M^2}{4},$$

con $M \equiv N$ (mód 2), $L \equiv \mp 1$ (mód 3).

Observemos que la última expresión de p es única salvo el signo de L y M, pues en una expresión de este tipo es necesario que $L \equiv M$ (mód 2) para que p sea entero, es necesario que $L \not\equiv 0$ (mód 3) para que p no sea divisible entre 3 y, según el signo que elijamos para L, obtendremos necesariamente los valores correspondientes a la descomposición $p = \pi\bar{\pi}$ en factores primos congruentes con 1 o con -1 módulo 3. En particular, la expresión de p es única si exigimos que $L, M > 0$. Como vamos a usar a menudo estos hechos, conviene recogerlos en un enunciado:

Teorema 15.4 *Todo primo $p \equiv 1$ (mód 3) se puede expresar en la forma*

$$p = \frac{L^2 + 27M^2}{4},$$

donde L y M están unívocamente determinados salvo el signo. Fijado un signo para L, tenemos que $p = \pi\bar{\pi}$, donde

$$\pi = \frac{L + 3M\sqrt{-3}}{2}, \qquad \bar{\pi} = \frac{L - 3M\sqrt{-3}}{2}\zeta$$

son primos de Eisenstein primarios (de modo que si $L \equiv \pm 1$ (mód 3), entonces $\pi \equiv \mp 1$ (mód 3)).

El resto cúbico de 2 Retomemos el problema de caracterizar los primos de la forma $p \equiv 1$ (mód 3) respecto a los cuales 2 es un resto cúbico. Lo que observó Euler es que existe un criterio muy simple en términos de la expresión de p dada por el teorema anterior.

He aquí una tabla:

p	7	13	19	**31**	37	**43**	61	67	73	79	97	103
L	1	5	7	**4**	11	**8**	1	5	7	17	19	13
M	1	1	1	**2**	1	**2**	3	3	3	1	1	3
p	**109**	**127**	139	151	**157**	163	181	193	199	211	**223**	**229**
L	**2**	**20**	23	19	**14**	25	7	23	11	13	**28**	**22**
M	**4**	**2**	1	3	**4**	1	5	3	5	5	**2**	**4**

Ahora la conjetura de Euler resulta patente: 2 es un resto cúbico módulo p si y sólo si L y M son pares, lo que a su vez equivale a que $p = a^2 + 27b^2$. ∎

Los restos cúbicos de 3 y 5 En la tabla siguiente señalamos los primos respecto a los cuales 3 es un resto cúbico:

p	7	13	19	31	37	43	**61**	**67**	**73**	79	97	**103**
L	1	5	7	4	11	8	**1**	**5**	**7**	17	19	**13**
M	1	1	1	2	1	2	**3**	**3**	**3**	1	1	**3**
p	109	127	139	**151**	157	163	181	**193**	199	211	223	229
L	2	20	23	**19**	14	25	7	**23**	11	13	28	22
M	4	2	1	**3**	4	1	5	**3**	5	5	2	4

Y en la siguiente señalamos los primos respecto a los que 5 es un resto cúbico:

p	7	**13**	19	31	37	43	61	**67**	73	79	97	103
L	1	**5**	7	4	11	8	1	**5**	7	17	19	13
M	1	**1**	1	2	1	2	3	**3**	3	1	1	3
p	109	**127**	139	151	157	**163**	**181**	193	**199**	**211**	223	229
L	2	**20**	23	19	14	**25**	**7**	23	**11**	**13**	28	22
M	4	**2**	1	3	4	**1**	**5**	3	**5**	**5**	2	4

Dejamos que el lector formule sus propias conjeturas. Más adelante las demostraremos a partir de la ley de reciprocidad cúbica. ∎

15.2 El símbolo cúbico

Pasemos ya a definir el símbolo cúbico análogo a los símbolos de Legendre y Jacobi para el caso cuadrático con el cual enunciaremos la ley de reciprocidad cúbica. Para entender la situación empezamos probando un análogo al criterio de Euler 5.8:

Teorema 15.5 *Sea π un primo de Eisenstein no asociado a λ. Entonces, un entero de Eisenstein α no divisible entre π es un resto cúbico módulo π si y sólo si*

$$\alpha^{(\mathrm{N}(\pi)-1)/3} \equiv 1 \ (\mathrm{mód}\ \pi).$$

DEMOSTRACIÓN: Sabemos que $\mathbb{Z}[\zeta]/(\pi)$ es un cuerpo[2] con $\mathrm{N}(\pi)$ elementos, luego su grupo de unidades U_π consta de $\mathrm{N}(\pi)-1$ clases (todas menos la clase nula). Ahora observamos que exactamente la tercera parte de dichas clases son restos cúbicos.

En efecto, basta considerar la aplicación $U_\pi \longrightarrow U_\pi$ dada por $u \mapsto u^3$. Se cumple que $u^3 = v^3$ si y sólo si $(v/u)^3 = 1$, es decir, si y sólo si $v/u = \bar{1}, \bar{\zeta}, \bar{\zeta}^2$, si y sólo si $v = u, \bar{\zeta}u, \bar{\zeta}^2 u$.

Notemos que las tres clases $\bar{1}, \bar{\zeta}, \bar{\zeta}^2$ son distintas, pues no puede ocurrir que π divida a la diferencia de dos de ellas, dado que

$$1-\zeta = \lambda, \quad 1-\zeta^2 = -(1+\zeta)\lambda = \zeta^2\lambda, \quad \zeta - \zeta^2 = \zeta\lambda$$

y estamos suponiendo que π no es asociado a λ. Por lo tanto, las clases de U_π se dividen en grupos de tres clases distintas con el mismo cubo, por lo que el número de cubos de U_π es la tercera parte de su número de elementos, es decir, es $(\mathrm{N}(\pi)-1)/3$.

Ahora consideramos el polinomio

$$x^{(\mathrm{N}(\pi)-1)/3} - 1.$$

Si α es un resto cúbico módulo π, es decir, si $\alpha \equiv \beta^3 \pmod{\pi}$, para cierto entero β, entonces

$$\alpha^{(\mathrm{N}(\pi)-1)/3} - 1 \equiv \beta^{\mathrm{N}(\pi)-1} - 1 \equiv 0 \pmod{\pi},$$

luego los $(\mathrm{N}(\pi)-1)/3$ restos cúbicos módulo π son raíces de este polinomio, y como tiene precisamente grado $(\mathrm{N}(\pi)-1)/3$, no puede tener más, por lo que concluimos que α es un resto cúbico módulo π si y sólo si es raíz del polinomio, es decir, si y sólo si $\alpha^{(\mathrm{N}(\pi)-1)/3} \equiv 1 \pmod{\pi}$. ∎

Incidentalmente, en la prueba del teorema anterior hemos visto que se cumple necesariamente $3 \mid \mathrm{N}(\pi)-1$, si bien esto puede verse directamente: si p es el primo racional divisible entre π, o bien $p \equiv 1 \pmod{3}$, en cuyo caso p se escinde en $\mathbb{Z}[\zeta]$, por lo que $\mathrm{N}(\pi) = p \equiv 1 \pmod{3}$, o bien $p \equiv -1 \pmod{3}$, en cuyo caso p se conserva, luego $\mathrm{N}(\pi) = p^2 \equiv 1 \pmod{3}$.

Ahora observamos que si α es un entero de Eisenstein no divisible entre π, el teorema 3.21 nos da que

$$\alpha^{\mathrm{N}(\pi)-1} \equiv 1 \pmod{\pi}.$$

(Esto es el equivalente al teorema de Fermat para enteros de Eisenstein.) Como el exponente es múltiplo de 3, podemos escribir equivalentemente:

$$(\alpha^{(\mathrm{N}(\pi)-1)/3})^3 \equiv 1 \pmod{\pi},$$

de modo que $\alpha^{(\mathrm{N}(\pi)-1)/3}$ es una raíz cúbica de la unidad en $\mathbb{Z}[\zeta]/(\pi)$, es decir, una raíz del polinomio x^3-1. Pero dicho polinomio sólo puede tener tres raíces en el cuerpo de restos, y dichas raíces son $\bar{1}, \bar{\zeta}, \bar{\zeta}^2$. Así pues:

[2]Por la observación previa al teorema 13.14.

Teorema 15.6 *Si π es un primo de Eisenstein que no divida a 3 y α es un entero de Eisenstein no divisible entre π, entonces existe un n tal que*

$$\alpha^{(\mathrm{N}(\pi)-1)/3} \equiv \zeta^n \pmod{\pi}.$$

Notemos que n es único módulo 3, pues ya hemos visto que las tres raíces de la unidad $1, \zeta, \zeta^2$ son distintas módulo π. Esto justifica la definición siguiente:

Definición 15.7 Si π es un primo de Eisenstein que no divida a 3 y α es un entero de Eisenstein arbitrario, definimos

$$\left(\frac{\alpha}{\pi}\right)_3 = \begin{cases} \zeta^n & \text{si } \pi \nmid \alpha \text{ y } \alpha^{(\mathrm{N}(\pi)-1)/3} \equiv \zeta^n \pmod{\pi}, \\ 0 & \text{si } \pi \mid \alpha. \end{cases}$$

De este modo se cumple que

$$\alpha^{(\mathrm{N}(\pi)-1)/3} \equiv \left(\frac{\alpha}{\pi}\right)_3 \pmod{\pi}.$$

También es obvio que

$$\left(\frac{\alpha\beta}{\pi}\right)_3 = \left(\frac{\alpha}{\pi}\right)_3 \left(\frac{\beta}{\pi}\right)_3,$$

así como que $(\alpha/\pi)_3$ sólo depende del resto de α módulo π. Más aún, es claro que

$$\overline{\left(\frac{\alpha}{\pi}\right)_3} = \left(\frac{\bar{\alpha}}{\bar{\pi}}\right)_3,$$

donde la barra denota la conjugación en $\mathbb{Z}[\zeta]$. El teorema 15.5 se expresa ahora así:

Teorema 15.8 *Si π es un primo de Eisenstein que no divida a 3 y α es un entero de Eisenstein no divisible entre π, entonces α es un resto cúbico módulo π si y sólo si $(\alpha/\pi)_3 = 1$.*

Ejemplo Consideremos de nuevo $\pi = \zeta - 3$, de modo que $\mathrm{N}(\pi) = 13$ y claramente $\zeta \equiv 3 \pmod{\pi}$. Calculamos

α	1	2	3	4	5	6	7	8	9	10	11	12
$\alpha^{(13-1)/3}$	1	3	3	9	1	9	9	1	9	3	3	1
$(\alpha/\pi)_3$	1	ζ	ζ	ζ^2	1	ζ^2	ζ^2	1	ζ^2	ζ	ζ	1

Por lo tanto:

$$\left(\frac{2}{\pi}\right)_3 \left(\frac{3}{\pi}\right)_3 = \zeta\zeta = \zeta^2 = \left(\frac{2\cdot 3}{\pi}\right)_3.$$

Esto explica por qué el producto $2 \cdot 3$ de dos restos no cúbicos da un resto no cúbico, mientras que $2 \cdot 7$ es un producto de restos no cúbicos que da un resto cúbico. ∎

Tenemos definida la versión cúbica del símbolo de Legendre, pero conviene definir también el análogo del símbolo de Jacobi:

Definición 15.9 Sean α, β son enteros de Eisenstein, donde β no es nulo ni unitario y sea $\beta = \pi_1 \cdots \pi_r$ la descomposición de β es factores primos. Si ninguno de ellos divide a 3, definimos

$$\left(\frac{\alpha}{\beta}\right)_3 = \prod_{j=1}^{r} \left(\frac{\alpha}{\pi_j}\right)_3.$$

Si β es una unidad definimos $(\alpha/\beta)_3 = 1$.

Claramente el símbolo de Jacobi es multiplicativo en sus dos argumentos y depende únicamente del resto de α módulo β. Además,

$$\overline{\left(\frac{\alpha}{\beta}\right)_3} = \left(\frac{\bar{\alpha}}{\bar{\beta}}\right)_3.$$

En particular, si m y n son enteros racionales y $3 \nmid n$, se cumple que $(m/n)_3 = 1$.

Ahora ya podemos enunciar la ley de reciprocidad:

Teorema 15.10 (Ley de reciprocidad cúbica) *Si α, β son enteros primarios de Eisenstein, entonces*

$$\left(\frac{\alpha}{\beta}\right)_3 = \left(\frac{\beta}{\alpha}\right)_3.$$

Leyes suplementarias *Si $\alpha = 1 + 3m + 3n\zeta$, entonces*

$$\begin{aligned}\left(\frac{\zeta}{\alpha}\right)_3 &= \zeta^{-m-n},\\ \left(\frac{\lambda}{\alpha}\right)_3 &= \zeta^{m}.\end{aligned}$$

En particular $(3/\alpha)_3 = \zeta^n$.

Demostración (de las leyes suplementarias): Supongamos en primer lugar que $\alpha = \pi$ es primo. Entonces

$$\mathrm{N}(\pi) = (1+3m)^2 + (3n)^2 - (1+3m)3n = 1 + 6m + 9m^2 + 9n^2 - 3n - 9mn,$$

luego

$$\frac{\mathrm{N}(\pi) - 1}{3} = 2m + 3m^2 + 3n^2 - n - 3mn \equiv -m - n \ (\text{mód } 3),$$

luego

$$\left(\frac{\zeta}{\pi}\right)_3 = \zeta^{(\mathrm{N}(\pi)-1)/3} = \zeta^{-m-n}.$$

Para probar el caso en que α es un entero de Eisenstein arbitrario (congruente con 1 módulo 3) observamos que si

$$\alpha = 1 + 3m + 3n\zeta, \qquad \beta = 1 + 3m' + 3n'\zeta,$$

una comprobación rutinaria muestra que

$$\alpha+\beta = 1+3m''+3n''\zeta,$$

donde $m'' \equiv m+m'$ (mód 3), $n'' \equiv n+n'$ (mód 3), por lo que, si el resultado vale para α y β, también vale para $\alpha\beta$, ya que

$$\left(\frac{\zeta}{\alpha\beta}\right)_3 = \left(\frac{\zeta}{\alpha}\right)_3\left(\frac{\zeta}{\beta}\right)_3 = \zeta^{-m-n}\zeta^{-m'-n'} = \zeta^{-m''-n''}.$$

Por lo tanto, como es válido para los divisores primos de α (que podemos elegir congruentes con 1 módulo 3), también vale para α. Si α es una unidad, necesariamente $\alpha = \pm 1$ (luego $m = n = 0$) y el resultado es trivial.

Observemos que el mismo argumento reduce la prueba de la segunda ley suplementaria al caso en que $\alpha = \pi$ es primo. Supongamos en primer lugar que $\pi = -q$, donde $q \equiv -1$ (mód 3) es un primo racional que se conserva primo en el anillo $\mathbb{Z}[\zeta]$. Entonces $-q = 1+3m$. Usamos que $3 = -\zeta^2\lambda^2$, luego $\lambda^2 = -3\zeta$ y

$$\left(\frac{\lambda}{\pi}\right)_3^2 = \left(\frac{-3\zeta}{\pi}\right)_3 = \left(\frac{-3}{-q}\right)_3\left(\frac{\zeta}{-q}\right)_3 = \zeta^{-m} = \zeta^{2m},$$

donde hemos usado que -3 es un resto cúbico módulo q y la primera ley suplementaria, ya probada. Elevando al cuadrado resulta

$$\left(\frac{\lambda}{\pi}\right)_3 = \zeta^m.$$

Ahora consideramos el caso en que $\pi = 1+3m+3n\zeta$ tiene norma prima p y vamos a probar la ley suplementaria usando la ley de reciprocidad cuadrática, que tenemos pendiente demostrar. Llamemos $a = 1+3m$, $b = 3n$, de modo que $\pi = a+b\zeta$. Entonces:

$$\left(\frac{a}{\pi}\right)_3 = \left(\frac{\pi}{a}\right)_3 = \left(\frac{a+b\zeta}{a}\right)_3 = \left(\frac{b\zeta}{a}\right)_3 = \left(\frac{\zeta}{a}\right)_3 = \zeta^{-m},$$

donde hemos usado que $(b/a)_3 = 1$ por la última observación tras 15.9.

Como

$$(a+b)\zeta = a+b\zeta-a+a\zeta = \pi - a(1-\zeta) \equiv -a\lambda \text{ (mód } \pi),$$

$$\left(\frac{a+b}{\pi}\right)_3 = \left(\frac{(a+b)\zeta\zeta^2}{\pi}\right)_3 = \left(\frac{-a\lambda\zeta^2}{\pi}\right)_3 = \left(\frac{-a}{\pi}\right)_3\left(\frac{\zeta}{\pi}\right)_3^2\left(\frac{\lambda}{\pi}\right)_3$$

$$= \zeta^{-m}(\zeta^{-m-n})^2\left(\frac{\lambda}{\pi}\right)_3 = \zeta^n\left(\frac{\lambda}{\pi}\right)_3,$$

luego

$$\left(\frac{\lambda}{\pi}\right)_3 = \zeta^{-n}\left(\frac{a+b}{\pi}\right)_3 = \zeta^{-n}\left(\frac{\pi}{a+b}\right)_3.$$

Pero $\pi = a + b\zeta \equiv -b + b\zeta = -b(1-\zeta) = -b\lambda$ (mód $a+b$), luego

$$\left(\frac{\pi}{a+b}\right)_3 = \left(\frac{-b\lambda}{a+b}\right)_3 = \left(\frac{\lambda}{a+b}\right)_3 = \left(\frac{\lambda^2}{a+b}\right)_3^2 = \left(\frac{-3\zeta}{a+b}\right)_3^2 = \left(\frac{\zeta}{a+b}\right)_3^2$$
$$= (\zeta^{-m-n})^2 = \zeta^{m+n},$$

luego

$$\left(\frac{\lambda}{\pi}\right)_3 = \zeta^{-n}\zeta^{m+n} = \zeta^m.$$

Por último demostramos la última afirmación tras la segunda ley complementaria:

$$\left(\frac{3}{\alpha}\right)_3 = \left(\frac{-\zeta^2\lambda^2}{\alpha}\right)_3 = \left(\frac{\zeta}{\alpha}\right)_3^2 \left(\frac{\lambda}{\alpha}\right)_3^2 = \zeta^{-2m-2n}\zeta^{2m} = \zeta^n. \qquad \blacksquare$$

Así pues, sólo queda pendiente demostrar la ley de reciprocidad cúbica propiamente dicha, y ésta se reduce al caso en que α y β son primos. En efecto:

Si α es una unidad, al ser primaria, tiene que ser $\alpha = \pm 1$, luego α es un cubo módulo todos los primos, luego

$$\left(\frac{\alpha}{\beta}\right)_3 = 1 = \left(\frac{\beta}{\alpha}\right)_3,$$

donde la segunda igualdad es por la definición del símbolo de Jacobi cúbico. Lo mismo vale si β es una unidad, así que podemos suponer que ninguno de los dos es unitario. Teniendo en cuenta que, por 15.3, ambos se descomponen en productos de primos primarios, si tenemos probada la ley de reciprocidad para primos, sólo tenemos que descomponer α y β en factores primos primarios, descomponer $(\alpha/\beta)_3$ en los productos correspondientes, invertir cada factor mediante la ley de reciprocidad para primos y luego volver a agruparlos, para obtener $(\beta/\alpha)_3$.

15.3 Aplicaciones de la reciprocidad cúbica

Probaremos la ley de reciprocidad cúbica en la sección 15.5, pero antes vamos a mostrar algunas de sus consecuencias. Empecemos por su uso más básico para determinar el carácter cúbico de un número dado módulo un primo dado:

Ejemplo Vamos a determinar el carácter cúbico de 17 módulo 31.

Como $31 \equiv 1$ (mód 3) el problema no es trivial. Un divisor primo de 31 en $\mathbb{Z}[\zeta]$ es $\pi = \zeta - 5$. Observemos que $\zeta - 5 \equiv \zeta + 1 = -\zeta^2$ (mód 3), luego, multiplicando por $-\zeta$, se cumple $-\zeta^2 + 5\zeta = 6\zeta + 1 \equiv 1$ (mód 3). Por otra parte, tenemos la división euclídea:

$$17 = -3(\zeta - 5) + 3\zeta + 2, \qquad \mathrm{N}(3\zeta + 2) = 7 < 31 = \mathrm{N}(\zeta - 5)$$

y además $3\zeta+2 \equiv -1 \pmod 3$, luego

$$\left(\frac{17}{\zeta-5}\right)_3 = \left(\frac{3\zeta+2}{6\zeta+1}\right)_3 = \left(\frac{6\zeta+1}{3\zeta+2}\right)_3 = \left(\frac{\zeta}{3\zeta+2}\right)_3 = \left(\frac{\zeta}{-3\zeta-2}\right)_3$$

$$= \left(\frac{\zeta}{1-3-3\zeta}\right)_3 = \zeta^{1+1} = \zeta^2,$$

donde hemos dividido de nuevo

$$6\zeta+1 = (3\zeta+2)(\zeta+2)+\zeta, \qquad \mathrm{N}(\zeta) = 1 < 7 = \mathrm{N}(3\zeta+2).$$

Concluimos que 17 no es un resto cúbico módulo 31. ∎

Ejemplo Vamos a calcular el carácter cúbico de 2 módulo 31.

Como antes:

$$\left(\frac{2}{\zeta-5}\right)_3 = \left(\frac{2}{6\zeta+1}\right)_3 = \left(\frac{6\zeta+1}{2}\right)_3 = \left(\frac{1}{2}\right)_3 = 1.$$

Por lo tanto 2 es un resto cúbico módulo 31. ∎

Seguidamente demostraremos los resultados que hemos discutido en la primera sección de este capítulo:

Ejemplo 2 *es un resto cúbico módulo un primo* $p \equiv 1 \pmod 3$ *si y sólo si éste es de la forma* $p = a^2 + 27b^2$.

En efecto, factorizamos $p = \pi\pi'$. Multiplicando por una unidad podemos exigir que $\pi \equiv 1 \pmod 3$. Entonces 2 es un resto cúbico módulo p si y sólo si $(2/\pi) = 1$. Por la ley de reciprocidad y la definición del símbolo cúbico (teniendo en cuenta que $(\mathrm{N}(2)-1)/3 = 1$):

$$\left(\frac{2}{\pi}\right)_3 = \left(\frac{\pi}{2}\right)_3 \equiv \pi \pmod 2,$$

luego 2 es un resto cúbico módulo p si y sólo si p es divisible entre un primo $\pi \equiv 1 \pmod 3$, $\pi \equiv 1 \pmod 2$.

En tal caso, $\pi = 6u+1+6v\zeta$ o, equivalentemente, $\pi = 6u+1-3v+3v\sqrt{-3}$. Haciendo $a = 6u+1-3v$, $b = v$, concluimos que $p = \mathrm{N}(\pi) = a^2 + 27b^2$.

Recíprocamente, si $p = a^2 + 27b^2$, ciertamente $3 \nmid a$ y, eligiendo el signo, podemos exigir que $a \equiv 1 \pmod 3$. Entonces $\pi = a+3b+6b\zeta$ cumple $\mathrm{N}(\pi) = p$, $\pi \equiv a \equiv 1 \pmod 3$, $\pi \equiv a+b \equiv 1 \pmod 2$ (si a y b fueran ambos pares o ambos impares p sería par). ∎

Ejemplo 3 *es un resto cúbico módulo un primo* $p \equiv 1 \pmod 3$ *si y sólo si su expresión en la forma* $p = (L^2 + 27M^2)/4$ *con* $L, M > 0$ *cumple que* $3 \mid M$.

En efecto, sea π un divisor primo de p en k, que podemos tomar de modo que $\pi \equiv 1 \pmod 3$, y entonces será de la forma $\pi = 1 + 3m + 3n\zeta$. Se cumplirá que 3 es un resto cúbico módulo p si y sólo si

$$\left(\frac{3}{\pi}\right)_3 = \zeta^n = 1,$$

es decir, si y sólo si $3 \mid n$, pero la expresión en términos de L y M se obtiene haciendo $M = n$, por lo que la condición equivale a que $3 \mid M$.

Es interesante señalar también que 3 es un resto cúbico módulo p si y sólo si $\pi = x + 9y\zeta$, lo que a su vez equivale a que $p = x^2 - 9xy + 81y^2$ o, pasando a una forma reducida, si y sólo si $p = x^2 + xy + 61y^2$. Con estas formas cuadráticas de discriminante -243 sucede lo mismo que con las de discriminante -108: hay tres clases de equivalentcia estricta, o dos si consideramos la equivalencia no estricta, que admiten como representantes las formas

$$x^2 + xy + 61y^2, \qquad 7x^2 + 3xy + 9y^2,$$

y un primo p está representado por una de ellas si y sólo si $p \equiv 1 \pmod 3$, pero todas las formas de dicho discriminante son del mismo género, por lo que la teoría de géneros no nos permite distinguir qué primos son de una forma y cuáles de la otra. ∎

Ahora probamos un resultado general:

Teorema 15.11 *Sea p un primo de la forma*

$$p = \frac{L^2 + 27M^2}{4},$$

donde el signo de L se elige para que $L \equiv -1 \pmod 3$. Sea $q > 2$ un primo distinto de p, sea

$$n = \frac{1}{3}\left(q - \left(\frac{q}{3}\right)\right)$$

y

$$g_n = \sum_{\substack{1 \le j \le n \\ \text{impar}}} \binom{n}{j} (-1)^{(j-1)/2}\, 3^{(3j-1)/2} L^{n-j} M^j.$$

Entonces, q es un resto cúbico módulo p si y sólo si $q \mid g_n$.

Antes de probar el teorema vamos a ver que —aunque parezca aparatoso— sus primeros casos particulares son muy sencillos. En general, notemos que $n = (q \pm 1)/3$, donde el signo es el necesario para que sea entero. Por completitud incluimos también los dos casos probados previamente para $q = 2, 3$:

1. 2 *es un resto cúbico módulo p si y sólo si $L \equiv M \equiv 0 \pmod 2$.*
2. 3 *es un resto cúbico módulo p si y sólo si $M \equiv 0 \pmod 3$.*

3. 5 *es un resto cúbico módulo p si y sólo si* $LM \equiv 0 \pmod 5$.
   En este caso $n = 2$ y $g_2 = 18LM$, luego $5 \mid g_5$ equivale a $5 \mid LM$.
4. 7 *es un resto cúbico módulo p si y sólo si* $LM \equiv 0 \pmod 7$.
   Ahora también $n = 2$, luego la conclusión es la misma.
5. 11 *es un resto cúbico módulo p si y sólo si*
   $$LM(L - 3M)(L + 3M) \equiv 0 \pmod{11}.$$
   Ahora hemos de considerar
   $$g_4 = 36L^3M - 324LM^3 = 36LM(L^2 - 9M^2) = 36LM(L-3M)(L+3M).$$
6. 13 *es un resto cúbico módulo p si y sólo si*
   $$LM(L - 3M)(L + 3M) \equiv 0 \pmod{13}.$$
   Nuevamente $n = 4$.

DEMOSTRACIÓN: Sea $\pi = (L + 3M\sqrt{-3})/2$, donde podemos suponer que $L \equiv -1 \pmod 3$. Así $p = \pi\bar{\pi}$. Supongamos en primer lugar que $q \equiv 1 \pmod 3$. En tal caso q se escinde en $\mathbb{Z}[\zeta]$, en la forma $q = \rho\bar{\rho}$, para cierto primo ρ, que podemos tomar $\rho \equiv 1 \pmod 3$.

Como $\mathbb{Z}[\zeta]/(\pi) \cong \mathbb{Z}/p\mathbb{Z}$, tenemos que q es un resto cúbico módulo p si y sólo si $(q/\pi)_3 = 1$, lo que, por la ley de reciprocidad, equivale a que $(\pi/q)_3 = 1$, pero

$$\left(\frac{\pi}{q}\right)_3 = \left(\frac{\pi}{\rho}\right)_3 \left(\frac{\pi}{\bar{\rho}}\right)_3 = \left(\frac{\pi}{\rho}\right)_3 \left(\frac{\bar{\pi}}{\rho}\right)_3^{-1}.$$

Por lo tanto, concluimos que q es un resto cúbico módulo p si y sólo si

$$\left(\frac{\pi}{\rho}\right)_3 = \left(\frac{\bar{\pi}}{\rho}\right)_3,$$

lo que, por definición del símbolo cúbico equivale a que

$$\pi^n \equiv \bar{\pi}^n \pmod{\rho}.$$

Veamos que llegamos a esta misma caracterización si suponemos que $q \equiv -1 \pmod 3$. Como en el caso anterior, q es un resto cúbico módulo p si y sólo si $(\pi/q) = 1$, pero ahora q es primo en $\mathbb{Z}[\zeta]$, luego, por definición del símbolo cúbico, esto equivale a que $\pi^{(q^2-1)/3} \equiv 1 \pmod q$ o también a que

$$(\pi^{q-1})^n \equiv 1 \pmod q.$$

Ahora observamos que, en general, si $\alpha = a + b\zeta$, entonces

$$\alpha^q \equiv a^q + b^q\zeta^q \equiv a + b\zeta^{-1} \equiv \bar{\alpha} \pmod q.$$

En particular, $\pi^{q-1} \equiv \bar{\pi}/\pi \pmod q$, luego llegamos igualmente a la equivalencia

$$\pi^n \equiv \bar{\pi}^n \pmod q.$$

De este modo, podemos tratar conjuntamente los dos casos sin más que considerar que ρ es un factor primo de q (que será $\rho = q$ en el segundo caso). La condición a la que hemos llegado es

$$\frac{(L+3M\sqrt{-3})^n}{2^n} \equiv \frac{(L-3M\sqrt{-3})^n}{2^n} \text{ (mód } \rho).$$

Como $\rho \nmid 2$, tenemos que 2 es una unidad módulo ρ, por lo que esto equivale a

$$(L+3M\sqrt{-3})^n \equiv (L-3M\sqrt{-3})^n \text{ (mód } \rho),$$

o también a

$$\sum_j \binom{n}{j} L^{n-j} 3^j M^j (\sqrt{-3})^j \equiv \sum_j \binom{n}{j} L^{n-j} (-3)^j M^j (\sqrt{-3})^j \text{ (mód } \rho),$$

pero los términos para j par se cancelan y queda $g_n\sqrt{-3} \equiv 0$ (mód ρ).

Como $\sqrt{-3} \mid 3$ y $\rho \nmid 3$, esto equivale a que $\rho \mid g_n$ y, como g_n es un entero racional, en el caso en que $q \equiv 1$ (mód 3) (y trivialmente en el caso opuesto), esto equivale a $q \mid g_n$. ∎

La prueba del teorema anterior muestra también lo siguiente:

Teorema 15.12 *Si p es un primo de la forma*

$$p = \frac{L^2+27M^2}{4},$$

todos los divisores de LM son restos cúbicos módulo p.

Demostración: Basta probar que todo primo $q \mid LM$ tal que $q \equiv 1$ (mód 3) es un resto cúbico módulo p. La prueba del teorema anterior muestra que q es un resto cúbico módulo p si y sólo si

$$\left(\frac{\pi}{\rho}\right)_3 = \left(\frac{\bar{\pi}}{\rho}\right)_3,$$

o, equivalentemente,

$$\left(\frac{L+3M\sqrt{-3}}{\rho}\right)_3 = \left(\frac{L-3M\sqrt{-3}}{\rho}\right)_3.$$

Ahora bien, según si $q \mid L$ o $q \mid M$ esto equivale a una de las dos igualdades

$$\left(\frac{3M\sqrt{-3}}{\rho}\right)_3 = \left(\frac{-3M\sqrt{-3}}{\rho}\right)_3, \qquad \left(\frac{L}{\rho}\right)_3 = \left(\frac{L}{\rho}\right)_3,$$

y las dos se cumplen trivialmente. ∎

Observemos[3] que 4 es un resto cúbico módulo un primo p si y sólo si lo es 2, luego el primer número no primo para el que no es trivial el problema de módulo qué primos es un resto cúbico es $n = 6$. Euler conjeturó la respuesta:

[3]En general, si un cuerpo k contiene una raíz $\alpha = \sqrt[3]{u^2}$, entonces $\alpha^2/u = \sqrt[3]{u}$.

Ejemplo *Dado un primo $p = a^2 + 3b^2$, se cumple que 6 es un resto cúbico módulo p si y sólo si $9 \mid b$ o $9 \mid a \pm 2b$.*

Esto puede probarse mediante una laboriosa aplicación sistemática de la reciprocidad cúbica. El problema principal con que nos encontramos es que si llamamos $\pi = a+b\sqrt{-3}$, no es necesariamente cierto que $\pi \equiv \pm 1$ (mód 3), luego tenemos que multiplicarlo por una unidad adecuada para aplicar la reciprocidad. Explícitamente, la tabla siguiente nos da la unidad congruente con π módulo 3 en función de los restos módulo 3 de a y b. Notemos que a no puede ser múltiplo de 3 o π sería divisible entre el primo $\sqrt{-3}$:

	0	1	2
1	1	ζ^2	ζ
2	-1	$-\zeta$	$-\zeta^2$

Para calcular la tabla observamos que $\sqrt{-3} = 2\zeta + 1$, luego $\pi = a + b + 2b\zeta$. Por ejemplo, en el caso $a \equiv b \equiv 2$ (mód 3), tenemos que

$$\pi = a + b\sqrt{-3} \equiv 2 + 2\sqrt{-3} = 4 + 4\zeta \equiv 1 + \zeta = -\zeta^2 \text{ (mód 3)}.$$

Alternativamente, la unidad ϵ que hace que $\epsilon\pi \equiv 1$ (mód 3) es

ϵ	0	1	2
1	1	ζ	ζ^2
2	-1	$-\zeta^2$	$-\zeta$

Para calcular $(3/\pi)_3$ calculamos $(3/\epsilon\pi)_3 = \zeta^n$ según las leyes suplementarias:

$(3/\pi)_3$	0	1	2
1	$\zeta^{-b/3}$	$\zeta^{(a-b)/3}$	$\zeta^{-(a+b)/3}$
2	$\zeta^{b/3}$	$\zeta^{(a+b)/3}$	$\zeta^{(b-a)/3}$

Por ejemplo, en el caso $a \equiv b \equiv 1$ (mód 3) tenemos que $\epsilon = \zeta$ y

$$\epsilon\pi = (a+b)\zeta + 2b\zeta^2 = -2b + (a-b)\zeta,$$

luego $n = (a-b)/3$. Para calcular $(2/\pi)_3 = (2/\epsilon\pi)_3 = (\epsilon\pi/2)_3$ necesitamos conocer los restos de a y b módulo 2, luego en total tenemos que considerar los restos módulo 6. Observemos que la expresión $\pi = a+b+2b\zeta$ requiere que $a+b$ sea impar, lo que explica los huecos en la tabla siguiente:

$(2/\pi)_3$	0	1	2	3	4	5
1	1	–	ζ^2	–	ζ	–
2	–	ζ^2	–	1	–	ζ
4	–	ζ	–	1	–	ζ^2
5	1	–	ζ	–	ζ^2	–

Por ejemplo, si $a \equiv 4$ (mód 6), $b \equiv 5$ (mód 6), tenemos que $\epsilon = \zeta^2$ y

$$\left(\frac{2}{\pi}\right)_3 = \left(\frac{\zeta^2\pi}{2}\right)_3 = \left(\frac{(a+b)\zeta^2 + 2b}{2}\right)_3 = \left(\frac{b - a - (a+b)\zeta}{2}\right)_3$$

$$= \left(\frac{1+\zeta}{2}\right)_3 = \left(\frac{-\zeta^2}{2}\right)_3 = \left(\frac{\zeta}{-2}\right)_3^2 = \zeta^2,$$

donde hemos aplicado la ley suplementaria a $-2 = 1 + 3(-1) + 0\zeta$.

Tenemos que 6 será un resto cúbico módulo p si y sólo si

$$\left(\frac{6}{\pi}\right)_3 = \left(\frac{2}{\pi}\right)_3 \left(\frac{3}{\pi}\right)_3 = 1.$$

Supongamos en primer lugar que $b \equiv 0$ (mód 3). En este caso

$$a \pm 2b \equiv a \not\equiv 0 \text{ (mód 3)}$$

luego tenemos que probar que $(6/\pi) = 1$ si y sólo si $9 \mid b$. Por otra parte, la tabla de $(2/\pi)$ muestra que $(2/\pi) = 1$, luego $(6/\pi) = (3/\pi)$. La tabla de $(3/\pi)$ muestra que este símbolo vale 1 si y sólo si $3 \mid b/3$, es decir, si y sólo si $9 \mid b$, como había que probar.

Supongamos ahora que $b \equiv \pm 1$ (mód 3), en cuyo caso no puede suceder que $9 \mid b$, luego tenemos que probar que $(6/\pi)_3 = 1$ si y sólo si $9 \mid a \pm 2b$. Combinando las dos tablas obtenemos:

$(6/\pi)_3$	1	2	4	5
1	–	$\zeta^{2-(a+b)/3}$	$\zeta^{1+(a-b)/3}$	–
2	$\zeta^{2+(a+b)/3}$	–	–	$\zeta^{1+(b-a)/3}$
4	$\zeta^{1+(a-b)/3}$	–	–	$\zeta^{2-(a+b)/3}$
5	–	$\zeta^{1+(b-a)/3}$	$\zeta^{2+(a+b)/3}$	–

Cada entrada de esta tabla da lugar a 9 subcasos según los restos de a y b módulo 18. Por ejemplo, el caso $a \equiv 2$ (mód 6), $b \equiv 1$ (mód 6) da lugar a los casos

	1	7	13
2	1	ζ^2	ζ
8	ζ^2	ζ	1
14	ζ	1	ζ^2

Podemos comprobar que los unos se encuentran justamente en los casos que cumplen $9 \mid a \pm 2b$. Igualmente se comprueban los casos correspondientes a las entradas restantes de la tabla. ∎

15.4 Sumas de Jacobi

Recordemos que en 9.21 hemos definido la suma de Gauss asociada a un carácter modular χ. Aquí necesitamos considerar únicamente caracteres módulo un primo impar p, de modo que

$$G_a(\chi) = \sum_{r \in \mathbb{Z}_p} \chi(r)\omega^{ar},$$

donde $\omega = e^{2a\pi i/p}$. Notemos que tiene sentido la expresión ω^{ar} porque las potencias de ω dependen sólo del resto módulo p del exponente.

Observemos que todos los caracteres módulo p son primitivos (definición 9.18) salvo el carácter principal χ_0, dado por $\chi_0(x) = 1$ para todo $x \in U_p$, pues si un carácter no es primitivo tiene que ser inducido por un carácter módulo un divisor de p, y la única posibilidad es que se trate del único carácter módulo 1.

Usaremos en varias ocasiones el teorema [ITAn 7.22], según el cual, si χ es un carácter módulo p, entonces

$$\sum_{k\in\mathbb{Z}_p} \chi(k) = \begin{cases} p-1 & \text{si } \chi = \chi_0, \\ 0 & \text{si } \chi \neq \chi_0. \end{cases}$$

He aquí un último hecho elemental:

$$\overline{G(\chi)} = \chi(-1)G(\bar{\chi}). \tag{15.1}$$

En efecto, tenemos que

$$\overline{G(\chi)} = \sum_{t\in\mathbb{Z}_p} \bar{\chi}(t)\omega^{-t}.$$

Ahora usamos que $-\bar{0}, \ldots, -\overline{p-1}$ recorren todas las clases de $\mathbb{Z}_p$, luego podemos cambiar t por $-t$ en la suma anterior:

$$\overline{G(\chi)} = \sum_{t\in\mathbb{Z}_p} \bar{\chi}(-t)\omega^{t} = \bar{\chi}(-1) \sum_{t\in\mathbb{Z}_p} \bar{\chi}(t)\omega^{t} = \chi(-1)G(\bar{\chi}),$$

donde hemos usado que $\bar{\chi}(-1) = \chi(-1)$, ya que, como $(-1)^2 = 1$, necesariamente $\chi(-1)^2 = 1$, luego $\chi(-1) = \pm 1$ es un número real, luego coincide con su conjugado. ∎

Las sumas de Jacobi que introducimos ahora nos darán una expresión para el producto de las sumas de Gauss de dos caracteres módulo un mismo primo p:

Definición 15.13 Si χ y ψ son dos caracteres módulo p, definimos la *suma de Jacobi*

$$J(\chi,\psi) = \sum_{t\in\mathbb{Z}_p} \chi(t)\psi(1-t).$$

Notemos que $J(\chi,\psi) = J(\psi,\chi)$.

Teorema 15.14 *Sean χ y ψ dos caracteres no principales módulo p. Entonces:*

1. $J(\chi,\chi^{-1}) = -\chi(-1)$,
2. *Si $\psi \neq \chi^{-1}$, entonces*

$$J(\chi,\psi) = \frac{G(\chi)G(\psi)}{G(\chi\psi)}.$$

DEMOSTRACIÓN: Para probar 1) observamos que

$$J(\chi,\chi^{-1}) = \sum_{t\in\mathbb{Z}_p} \chi(t)\chi^{-1}(1-t) = \sum_{t\in\mathbb{Z}_p\setminus\{1\}} \chi(t(1-t)^{-1})$$

Si llamamos $c = t(1-t)^{-1}$, tenemos que $(1-t)c = t$, o también $(c+1)t = c$, luego si $c \neq -1$ podemos despejar $t = c(c+1)^{-1}$. Esto significa que, cuando t recorre $\mathbb{Z}_p \setminus \{1\}$, la expresión $t(1-t)^{-1}$ recorre $\mathbb{Z}_p \setminus \{-1\}$, luego

$$J(\chi,\chi^{-1}) = \sum_{c\in\mathbb{Z}_p\setminus\{-1\}} \chi(c) = \sum_{c\in\mathbb{Z}_p} \chi(c) - \chi(-1) = -\chi(-1),$$

donde hemos usado [ITAn 7.22]. Ahora probamos 2):

$$G(\chi)G(\psi) = \Big(\sum_{s\in\mathbb{Z}_p} \chi(s)\omega^s\Big)\Big(\sum_{t\in\mathbb{Z}_p} \psi(s)\omega^t\Big) = \sum_{s,t\in\mathbb{Z}_p} \chi(s)\psi(t)\omega^{s+t}$$

$$= \sum_{x\in\mathbb{Z}_p} \Big(\sum_{s\in\mathbb{Z}_p} \chi(s)\psi(\bar{x}-s)\Big)\omega^x.$$

Si $x = 0$, tenemos que

$$\sum_{s\in\mathbb{Z}_p} \chi(s)\psi(-s) = \psi(-1)\sum_{s\in\mathbb{Z}_p} (\chi\psi)(s) = 0,$$

por [ITAn 7.22] Si $x > 0$, entonces, cuando t recorre $\mathbb{Z}_p$, lo mismo sucede con $s = \bar{x}t$, luego

$$\sum_{s\in\mathbb{Z}_p} \chi(s)\psi(\bar{x}-s) = \sum_{t\in\mathbb{Z}_p} \chi(\bar{x}t)\psi(\bar{x}-\bar{x}t) =$$

$$\chi(x)\psi(x)\sum_{t\in\mathbb{Z}_p} \chi(t)\psi(1-t) = \chi(x)\psi(x)J(\chi,\psi).$$

Por lo tanto:

$$G(\chi)G(\psi) = \sum_{x\in\mathbb{Z}_p} (\chi\psi)(x)J(\chi,\psi) = G(\chi\psi)J(\chi,\psi).$$

∎

Vamos a extraer algunas consecuencias de este teorema. En primer lugar, combinándolo con el teorema 9.23 obtenemos:

Teorema 15.15 *Si χ, ψ y $\chi\psi$ son caracteres no principales módulo p, entonces*

$$|J(\chi,\psi)| = \sqrt{p}.$$

Teorema 15.16 *Si χ es un carácter módulo p de orden $n > 2$, entonces*

$$G(\chi)^n = \chi(-1)pJ(\chi,\chi)J(\chi,\chi^2)\cdots J(\chi,\chi^{n-2}).$$

DEMOSTRACIÓN: Como $\chi \neq \chi^{-1}$, el teorema anterior nos da que

$$G(\chi)^2 = J(\chi,\chi)G(\chi^2).$$

Por lo tanto:

$$G(\chi)^3 = J(\chi,\chi)G(\chi^2)G(\chi).$$

Si $n > 3$, como $\chi^2 \neq \chi^{-1}$, aplicando de nuevo el teorema obtenemos que

$$G(\chi)^3 = J(\chi,\chi)J(\chi,\chi^2)G(\chi^3).$$

Procediendo de este modo, llegamos a que

$$G(\chi)^{n-1} = J(\chi,\chi)J(\chi,\chi^2)\cdots J(\chi,\chi^{n-2})G(\chi^{n-1}),$$

luego

$$G(\chi)^n = J(\chi,\chi)J(\chi,\chi^2)\cdots J(\chi,\chi^{n-2})G(\chi^{n-1})G(\chi).$$

Ahora, como $\chi^{n-1} = \chi^{-1}$, el teorema 9.23 y (15.1) nos dan que

$$G(\chi^{n-1})G(\chi) = G(\bar{\chi})G(\chi) = \chi(-1)\overline{G(\chi)}G(\chi) = \chi(-1)|G(\chi)|^2 = \chi(-1)p,$$

y esto nos da la fórmula del enunciado. ∎

En particular, si χ tiene orden 3, tenemos $\chi(-1) = \chi((-1)^3) = \chi^3(-1) = 1$, luego la fórmula del teorema anterior se reduce a

$$G(\chi)^3 = pJ(\chi,\chi).$$

Más aún, todos los valores que toma χ son potencias de $\zeta = \omega_3$, por lo que $J(\chi,\chi) \in \mathbb{Z}[\zeta]$ es un entero de Eisenstein, luego, por 15.15,

$$\mathrm{N}(J(\chi,\chi)) = |J(\chi,\chi)|^2 = p$$

luego $J(\chi,\chi)$ es un primo de Eisenstein divisor de p. Vamos a probar que es primario. Más precisamente, vamos a ver que $J(\chi,\chi) \equiv -1$ (mód 3):

Teorema 15.17 *Si χ es un carácter cúbico módulo un primo $p \equiv 1$ (mód 3), entonces $J(\chi,\chi) = a + b\zeta$, donde $3 \mid b$ y $a \equiv -1$ (mód 3).*

Demostración: Sea $\omega = e^{2\pi i/p}$ y $\Omega = e^{2\pi i/3p}$, de modo que $\omega = \Omega^3$ y $\zeta = \Omega^p$. Vamos a trabajar en el anillo $\mathbb{Z}[\Omega]$ formado por todos los números complejos de la forma $f(\Omega)$, donde $f(x)$ es un polinomio con coeficientes enteros. El único hecho no trivial que necesitamos sobre este anillo es que, como Ω es un entero algebraico (porque es raíz del polinomio $x^{3p}-1$), todos los elementos de $\mathbb{Z}[\Omega]$ son enteros algebraicos (por 8.15).

Vamos a considerar congruencias módulo 3 en el anillo $\mathbb{Z}[\Omega]$. Como el cociente $\mathbb{Z}[\Omega]/(3)$ es un anillo de característica 3, podemos aplicar el teorema 3.16:

$$G(\chi)^3 \equiv \Big(\sum_{t\in\mathbb{Z}_p} \chi(t)\omega^t\Big)^3 \equiv \sum_{t\in\mathbb{Z}_p} \chi^3(t)\omega^{3t} = \sum_{t=0}^{p-1} \omega^{3t} - 1$$

$$= \frac{\omega^{3p}-1}{\omega^3-1} - 1 = -1 \text{ (mód 3)}.$$

Por lo tanto, teniendo en cuenta que $p \equiv 1$ (mód 3),

$$a + b\zeta = J(\chi,\chi) \equiv pJ(\chi,\chi) = G(\chi)^3 \equiv -1 \text{ (mód 3)}.$$

Por otro lado, es obvio que $J(\bar{\chi},\bar{\chi}) = \overline{J(\chi,\chi)} = a + b\bar{\zeta}$, y aplicando a $\bar{\chi}$ el razonamiento precedente vemos que $a + b\bar{\zeta} \equiv -1$ (mód 3), luego restando ambas congruencias llegamos a que $b(\zeta - \bar{\zeta}) \equiv 0$ (mód 3) o, equivalentemente, $3 \mid b(\zeta - \bar{\zeta})$, pero $\zeta - \bar{\zeta} = \sqrt{-3}$, luego $9 \mid -3b^2$, luego $3 \mid b^2$.

En este punto debemos recordar que estamos trabajando en $\mathbb{Z}[\Omega]$, de modo que lo que hemos probado es que 3 divide a b^2 en este anillo (no en $\mathbb{Z}$), pero en realidad es equivalente, pues lo que tenemos es que $b^3/3 \in \mathbb{Z}[\Omega]$, luego es a la vez un número racional y un entero algebraico, luego es un entero racional (teorema 8.18), luego $3 \mid b^2$ en $\mathbb{Z}$, luego $3 \mid b$, y esto implica a su vez que $3 \mid a+1$ en $\mathbb{Z}[\Omega]$, luego en $\mathbb{Z}$, y así concluimos que $a \equiv -1$ (mód 3). ∎

Con esto tenemos determinado el valor de $J(\chi,\chi)$ para un carácter cúbico:

Teorema 15.18 *Si $p = (L^2 + 27M^2)/4$ es primo, donde $L \equiv 1$ (mód 3), y χ es un carácter cúbico módulo p, entonces*

$$J(\chi,\chi) = \frac{L \pm 3M}{2} \pm 3M\zeta.$$

(El doble signo se debe a que hay dos caracteres cúbicos, y cada signo da lugar a la suma de Jacobi correspondiente.)

Demostración: Si $J(\chi,\chi) = a+b\zeta$, por la observación tras el teorema 15.16 sabemos que

$$\mathrm{N}(J(\chi,\chi)) = a^2 + b^2 - ab = p,$$

luego

$$4p = (2a-b)^2 + 3b^2 = L_0^2 + 27M_0^2,$$

donde llamamos $L_0 = 2a - b$, $M_0 = b/3$. El teorema anterior nos asegura que M_0 es entero, así como que $L_0 \equiv 1$ (mód 3). Sabemos que los valores L y M del enunciado están unívocamente determinados por p salvo signo, pero la condición $L_0 \equiv 1$ (mód 3) hace que $L_0 = L$, mientras que $M_0 = \pm M$. Despejando resulta que $a = (L \pm 3M)/2$, $b \pm 3M$. ∎

15.5 La prueba de la ley de reciprocidad

El último teorema previo que necesitamos para probar la ley de reciprocidad cúbica es el cálculo explícito de la suma de Jacobi para los caracteres definidos por el símbolo cúbico:

Teorema 15.19 *Sea π un primo primario de Eisenstein de norma prima p y sea χ_π el carácter módulo p dado por*

$$\chi_\pi(t) = \left(\frac{t}{\pi}\right)_3.$$

Entonces

$$J(\chi_\pi,\chi_\pi) = \begin{cases} \pi & \text{si } \pi \equiv -1 \text{ (mód 3)},\\ -\pi & \text{si } \pi \equiv 1 \text{ (mód 3)}. \end{cases}$$

DEMOSTRACIÓN: Vamos a probar que $\pi \mid J(\chi,\chi)$, con lo que $J(\chi,\chi)$ será uno de los asociados de π. Ahora bien, como $J(\chi,\chi) \equiv -1$ (mód 3), será necesariamente el indicado en el enunciado.

En efecto, por la definición del símbolo cúbico:

$$J(\chi_\pi,\chi_\pi) = \sum_{t\in\mathbb{Z}_p} \chi_\pi(t)\chi_\pi(1-t) \equiv \sum_{t=0}^{p-1} t^{(p-1)/3}(1-t)^{(p-1)/3}$$

$$= \sum_{t=0}^{p-1} t^{(p-1)/3} \sum_{j=0}^{(p-1)/3} \tbinom{(p-1)/3}{j}(-t)^j$$

$$= \sum_{j=0}^{(p-1)/3} \tbinom{(p-1)/3}{j}(-1)^j \sum_{t=0}^{p-1} t^{(p-1)/3+j} \pmod{\pi}.$$

Ahora observamos que $(p-1)/3 + j < p-1$, luego si u es una raíz primitiva módulo p, se cumple que $u^{(p-1)/3+j} \neq 1$, y la última suma queda invariante módulo p cuando se multiplica por esta clase de restos, luego tiene que ser nula módulo p. Concluimos que $J(\chi_\pi,\chi_\pi) \equiv 0 \pmod{\pi}$. ∎

Finalmente, vamos a demostrar la ley de reciprocidad cúbica. Según las observaciones posteriores al enunciado 15.10, es suficiente probarla cuando α y β son dos primos de Eisenstein distintos.

El caso en que $\alpha = p$, $\beta = q$ son primos racionales que se conservan en $\mathbb{Z}[\zeta]$ es trivial, pues entonces $p \equiv q \equiv -1$ (mód 3), luego p es un cubo módulo q y viceversa, es decir, tenemos que

$$\left(\frac{p}{q}\right)_3 = 1 = \left(\frac{q}{p}\right)_3.$$

Supongamos ahora que $\alpha = q \equiv -1$ (mód 3), mientras que $\beta = \pi$, donde $\mathrm{N}(\pi) = p \equiv 1$ (mód 3). Definimos

$$\chi_\pi(x) = \left(\frac{x}{\pi}\right)_3, \qquad \chi_q(\alpha) = \left(\frac{\alpha}{q}\right)_3.$$

Hemos probado que

$$G(\chi_\pi)^3 = pJ(\chi_\pi,\chi_p) = \pm p\pi,$$

luego, por definición del símbolo cúbico (trabajando de nuevo en el anillo $\mathbb{Z}[\Omega]$, como en la prueba del teorema 15.17):

$$G(\chi_\pi)^{q^2-1} = (p\pi)^{(q^2-1)/3} \equiv \chi_q(p\pi) \equiv \chi_q(\pi) \pmod{q},$$

donde hemos usado que $\chi_q(p) = 1$ porque p es un resto cúbico módulo q. Por lo tanto:

$$G(\chi_\pi)^{q^2} \equiv \chi_q(\pi)G(\chi_\pi) \pmod{q}.$$

Por otro lado, como el anillo $\mathbb{Z}[\Omega]/(q)$ tiene característica q,

$$G(\chi_\pi)^{q^2} \equiv \sum_{t=0}^{p-1} \chi_\pi^{q^2}(t)\omega^{q^2t} = \sum_{t=0}^{p-1} \chi_\pi(t)\omega^{q^2t} = G_{q^2}(\chi_\pi) = \bar{\chi}_\pi(q^2)G(\chi_\pi) \text{ (mód } q),$$

donde hemos usado que $q^2 \equiv 1$ (mód 3) y que $\chi_\pi(t)$ es una raíz cúbica de la unidad. Combinando las dos últimas congruencias, tenemos que

$$\bar{\chi}_\pi(q^2)G(\chi_\pi) \equiv \chi_q(\pi)G(\chi_\pi) \text{ (mód } q).$$

Pero el cuadrado de una raíz cúbica de la unidad es su inversa, luego en realidad $\bar{\chi}_\pi(q^2) = \chi_\pi(q)$ y así

$$\chi_\pi(q)G(\chi_\pi) \equiv \chi_q(\pi)G(\chi_\pi) \text{ (mód } q).$$

Por 9.23 sabemos que $G(\chi_\pi)\overline{G(\chi_\pi)} = p$, luego, multiplicando por la suma de Gauss conjugada,

$$\chi_\pi(q)p \equiv \chi_q(\pi)p \text{ (mód } q),$$

y multiplicando por el inverso de p módulo q, llegamos a que

$$\chi_\pi(q) \equiv \chi_q(\pi) \text{ (mód } q).$$

En principio, esta congruencia es en $\mathbb{Z}[\Omega]$, de modo que lo que hemos probado es que

$$\frac{\chi_\pi(q) - \chi_q(\pi)}{q} \in \mathbb{Z}[\Omega]$$

luego el cociente es un entero algebraico y también un elemento de $\mathbb{Q}(\zeta)$, luego es un elemento del anillo de enteros algebraicos de $\mathbb{Q}(\zeta)$, es decir, está en $\mathbb{Z}[\zeta]$, luego

$$\chi_\pi(q) \equiv \chi_q(\pi) \text{ (mód } q).$$

en $\mathbb{Z}[\zeta]$. Pero la unicidad del símbolo cúbico[4] implica que $\chi_\pi(q) = \chi_q(\pi)$, que es lo que afirma la ley de reciprocidad.

Supongamos, por último, que

$$\alpha = \pi, \quad \beta = \rho, \quad \mathrm{N}(\pi) = p, \quad \mathrm{N}(\rho) = q, \quad p \equiv q \equiv 1 \text{ (mód } 3).$$

Podemos suponer que $p \neq q$, pues si probamos este último caso cuando $p \neq q$, sólo queda la posibilidad de que $\alpha = \pi$, $\beta = \bar{\pi}$, y entonces

$$\begin{aligned}\left(\frac{\pi}{\bar{\pi}}\right)_3 &= \left(\frac{\pi+\bar{\pi}}{\bar{\pi}}\right)_3 = \left(\frac{\bar{\pi}}{\pi+\bar{\pi}}\right)_3 = \left(\frac{-\bar{\pi}}{\pi+\bar{\pi}}\right)_3\\ &= \left(\frac{\pi}{\pi+\bar{\pi}}\right)_3 = \left(\frac{\pi+\bar{\pi}}{\pi}\right)_3 = \left(\frac{\bar{\pi}}{\pi}\right)_3,\end{aligned}$$

donde usamos que $\bar{\pi} \nmid \pi + \bar{\pi}$, por lo que las dos aplicaciones de la ley de reciprocidad se reducen a los casos ya demostrados.

[4]En la prueba del teorema 15.5 hemos visto que dos raíces cúbicas de la unidad no pueden ser congruentes módulo un primo.

Como antes, consideramos $\omega = e^{2\pi i/p}$ y razonamos en el anillo $\mathbb{Z}[\Omega]$, donde $\Omega = \omega_{3p}$:

$$G(\chi_{\bar\pi})^{q-1} = (\pm p\bar\pi)^{(q-1)/3} \equiv \chi_\rho(p\bar\pi) \text{ (mód } \rho),$$

luego

$$G(\chi_{\bar\pi})^q \equiv \chi_\rho(p\bar\pi)G(\chi_{\bar\pi}) \text{ (mód } \rho).$$

Por otra parte, como $\rho \mid q$, el cociente $\mathbb{Z}[\Omega]/(\rho)$ tiene característica q, luego

$$G(\chi_{\bar\pi})^q \equiv \sum_{t\in\mathbb{Z}_p} \chi_{\bar\pi}(t)^q\omega^{qt} = \sum_{t\in\mathbb{Z}_p} \chi_{\bar\pi}(t)\omega^{qt} = G_q(\chi_{\bar\pi}) = \bar\chi_{\bar\pi}(q)G(\chi_{\bar\pi}) \text{ (mód } \rho).$$

Uniendo las dos congruencias resulta que

$$\chi_\rho(p\bar\pi)G(\chi_{\bar\pi}) \equiv \bar\chi_{\bar\pi}(q)G(\chi_{\bar\pi}) \text{ (mód } \rho).$$

Como en el caso precedente, multiplicando por el conjugado de la suma de Gauss queda

$$\chi_\rho(p\bar\pi)p \equiv \bar\chi_{\bar\pi}(q)p \text{ (mód } \rho),$$

y multiplicando por el inverso de p módulo q llegamos a que

$$\chi_\rho(p\bar\pi) \equiv \bar\chi_{\bar\pi}(q) \text{ (mód } \rho),$$

En principio esta congruencia es en $\mathbb{Z}[\Omega]$, pero como en el caso precedente concluimos que también vale en $\mathbb{Z}[\zeta]$, y de aquí pasamos a que

$$\chi_\rho(p\bar\pi) = \bar\chi_{\bar\pi}(q) = \chi_\pi(q),$$

por la última observación tras la definición 15.9. Multiplicando por $\chi_\pi(\rho)$ queda

$$\chi_\pi(\rho)\chi_\rho(p\bar\pi) = \chi_\pi(q\rho).$$

El mismo razonamiento cambiando $\bar\pi$ por ρ y ρ por π nos da que

$$\chi_\pi(q\rho) = \bar\chi_\rho(p) = \chi_\rho(p^2) = \chi_\rho(p\pi\bar\pi) = \chi_\rho(\pi)\chi_\rho(p\bar\pi),$$

donde hemos usado que el conjugado de una raíz cúbica de la unidad es su cuadrado. Por lo tanto

$$\chi_\pi(\rho)\chi_\rho(p\bar\pi) = \chi_\rho(\pi)\chi_\rho(p\bar\pi),$$

luego $\chi_\pi(\rho) = \chi_\rho(\pi)$, como había que probar. ∎

15.6 La congruencia $x^3 + y^3 \equiv 1$ (mód p)

El teorema 5.12 afirma en particular que si p es un primo impar, la congruencia

$$x^2 + y^2 \equiv 1 \text{ (mód } p)$$

tiene $p - (-1/p)$ soluciones módulo p. Los resultados que hemos probado sobre sumas de Jacobi nos permiten obtener un resultado similar para la congruencia

$$x^3 + y^3 \equiv 1 \text{ (mód } p).$$

La tabla siguiente muestra el número de soluciones N_p para los primeros primos:

p	2	3	5	**7**	11	**13**	17	**19**	23	29	**31**	**37**	41
N_p	2	3	5	**6**	11	**6**	17	**24**	23	29	**33**	**24**	41
p	**43**	47	53	59	**61**	**67**	71	**73**	**79**	83	89	**97**	101
N_p	**33**	47	53	59	**60**	**60**	71	**78**	**60**	83	89	**114**	101

Si $p = 3$ o $p \equiv -1 \pmod 3$ la situación es sencilla: como todos los números son cubos, la congruencia es equivalente a $x + y \equiv 1 \pmod p$, que obviamente tiene p soluciones. Sin embargo, cuando $p \equiv 1 \pmod p$, el patrón ya no está tan claro, y a pesar de ello Gauss encontró y demostró una fórmula general.

Teniendo en cuenta los resultados que hemos probado en este capítulo, no es descabellado tratar de encontrar la clave en la expresión de p en la forma $p = (L^2 + 27M^2)/4$. Dejamos aquí una tabla para que el lector ponga a prueba su perspicacia:

p	7	13	19	31	37	43	61	67	73	79	97
L	1	5	7	4	11	8	1	5	7	17	19
M	1	1	1	2	1	2	3	3	3	1	1
N_p	6	6	24	33	24	33	60	60	78	60	114

Vamos a abordar el problema para exponentes arbitrarios n, con la condición $n \mid p-1$. Así $d = (n, p-1) = n$ y el teorema 3.37 nos dice que, si $a \neq 0$, la ecuación $x^n = a$ tiene exactamente n soluciones en U_p o bien ninguna, mientras que si $a = 0$ tiene únicamente la solución $x = 0$. Esto puede expresarse así:

Teorema 15.20 *Si p es primo y $n \mid p-1$, el número de soluciones módulo p de la congruencia $x^n \equiv a \pmod p$ es*

$$\sum_{\chi^n = \chi_0} \chi(a),$$

donde la suma recorre los caracteres módulo p de orden divisor de n.

Demostración: Si $a \equiv 0 \pmod p$ hay una única solución módulo p, y por otro lado la suma vale 1, pues, de acuerdo con el convenio que hemos establecido, todos los sumandos son nulos excepto el correspondiente al carácter principal. A partir de aquí suponemos que $a \in U_p$.

Observemos que existen exactamente n caracteres que cumplen $\chi^n = \chi_0$. Esto se debe a que, fijada una raíz primitiva de la unidad u módulo p, cada carácter está determinado por $\chi(u)$, y el carácter cumplirá $\chi^n = \chi_0$ si y sólo si $\chi(u)^n = 1$, y hay exactamente n raíces n-simas de la unidad en $\mathbb{C}$, luego hay n caracteres posibles.

Si la congruencia tiene solución, el teorema 3.37 nos da que tiene exactamente n. Si una de ellas es $b^n = a$, entonces $\chi(a) = \chi(b)^n = \chi^n(b) = 1$, luego los n sumandos valen 1, y en efecto la suma es igual a n.

Por último, tenemos que probar que si la congruencia no tiene solución, entonces la suma vale 0. Para ello fijemos una raíz primitiva de la unidad u módulo p y sea ψ un carácter tal que $\psi(u)$ sea una raíz de la unidad de orden n. Así $\psi^n = \chi_0$ y si $a = u^b$, en la prueba del teorema 3.36 hemos visto que el hecho de que la congruencia no tenga solución equivale a que $n \nmid b$, por lo que $\psi(a) = \psi(u)^b \neq 1$.

Si $\chi_0, \ldots, \chi_n$ son todos los caracteres de orden divisor de n, es claro que $\psi\chi_0, \ldots, \psi\chi_n$ son esos mismos caracteres pero en otro orden, por lo que

$$\psi(a) \sum_{\chi^n=\chi_0} \chi(a) = \sum_{\chi^n=\chi_0} \chi(a),$$

luego

$$(\psi(a) - 1) \sum_{\chi^n=\chi_0} \chi(a) = 0,$$

de donde se sigue que la suma es nula. ∎

Como consecuencia:

Teorema 15.21 *Si p es un primo y n un entero tal que $p \equiv 1 \pmod{n}$, entonces el número de soluciones módulo p de la congruencia $x^n + y^n \equiv 1 \pmod{p}$ es igual a*

$$\sum_{\chi,\psi} J(\chi, \psi),$$

donde χ y ψ recorren los caracteres módulo p de orden divisor de n.

Demostración: Si, para cada $t \in \mathbb{Z}_p$, llamamos

$$N_t = \sum_{\chi^n=\chi_0} \chi(t),$$

por el teorema anterior sabemos que N_t es el número de soluciones de la ecuación $x^n = t$ en $\mathbb{Z}_p$, luego el número de soluciones módulo p de la congruencia del enunciado es

$$\sum_{t\in\mathbb{Z}_p} N_t N_{1-t} = \sum_{t\in\mathbb{Z}_p} \sum_{\chi^n=\chi_0} \chi(t) \sum_{\psi^n=\chi_0} \psi(1-t) = \sum_{\chi,\psi} J(\chi, \psi).$$

∎

Por ejemplo, para $n = 2$ hay un único carácter de orden 2 módulo p, que es el dado por $\chi(a) = (a/p)$, y el teorema anterior se reduce a que el número de soluciones de la congruencia $x^2 + y^2 \equiv 1 \pmod{p}$ es

$$J(\chi_0, \chi_0) + 2J(\chi_0, \chi) + J(\chi, \chi) = p - \chi(-1) = p - \left(\frac{-1}{p}\right),$$

donde hemos usado el teorema 15.14 para evaluar las sumas de Jacobi. Vemos que el resultado coincide con el dado por el teorema 5.12.

Para $n = 3$ tenemos dos caracteres cúbicos conjugados, digamos χ y $\bar{\chi}$, que cumplen $\chi(-1) = \chi((-1)^3) = \chi(-1)^3 = 1$, luego $G(\bar{\chi}) = \overline{G(\chi)}$, por (15.1). La definición de suma de Jacobi nos da entonces que $J(\bar{\chi}, \bar{\chi}) = \overline{J(\chi, \chi)}$.

Por lo tanto, el número de soluciones módulo p de la congruencia $x^3+y^3 \equiv 1 \pmod{p}$ es

$$J(\chi_0,\chi_0) + 2J(\chi,\bar{\chi}) + J(\chi,\chi) + J(\bar{\chi},\bar{\chi}) = p - 2\chi(-1) + J(\chi,\chi) + \overline{J(\chi,\chi)}$$

$$= p - 2 + 2\,\mathrm{Re}\,J(\chi,\chi).$$

Ahora sólo tenemos que evaluar la suma de Jacobi mediante 15.18:

Teorema 15.22 *Dado un primo $p = (L^2 + 27M^2)/3$, donde $L \equiv 1 \pmod{3}$, el número de soluciones módulo p de la congruencia $x^3 + y^3 \equiv 1 \pmod{p}$ es $p - 2 + L$.*

Demostración: Acabamos de probar que el número de soluciones viene dado por $p - 2 + 2\,\mathrm{Re}\,J(\chi,\chi)$, y el teorema 15.18 nos da que

$$J(\chi,\chi) = \frac{L \pm 3M}{2} \pm 3M\zeta = \frac{L \pm 3M}{2} \pm 3M\frac{-1+\sqrt{3}i}{2},$$

luego

$$\mathrm{Re}\,J(\chi,\chi) = \frac{L \pm 3M}{2} \pm \frac{-3M}{2} = \frac{L}{2},$$

de donde obtenemos la fórmula del enunciado. ∎

Copiamos de nuevo la tabla que habíamos presentado antes, pero ahora ajustamos el signo de L para que cumpla la condición $L \equiv 1 \pmod{3}$:

p	7	13	19	31	37	43	61	67	73	79	97
L	1	−5	7	−4	−11	−8	1	−5	7	−17	19
M	1	1	1	2	1	2	3	3	3	1	1
N_p	6	6	24	33	24	33	60	60	78	60	114

Capítulo XVI

La ley de reciprocidad bicuadrática

Gauss también conjeturó —aunque no pudo demostrar— una ley de reciprocidad bicuadrática, que, al igual que la ley de reciprocidad cúbica, sería demostrada por Eisenstein. Los hechos básicos sobre restos bicuadráticos módulo un primo p los tenemos ya demostrados, pero vamos a ilustrarlos con un par de ejemplos, para $p = 11, 13$:

$p = 11$

x	1	2	3	4	5	6	7	8	9	10
x^2	1	4	9	5	3	3	5	9	4	1
x^4	1	5	4	3	9	9	3	4	5	1

$p = 13$

x	1	2	3	4	5	6	7	8	9	10	11	12
x^2	1	4	9	3	12	10	10	12	3	9	4	1
x^4	1	3	3	9	1	9	9	1	9	3	3	1

Vemos que módulo 11 hay 5 restos bicuadráticos (la mitad de los 10 restos), que son los mismos que los restos cuadráticos. En cambio, módulo 13 hay 3 restos bicuadráticos (la cuarta parte de los 12 restos), que son la mitad de los 6 restos cuadráticos. Esto es consecuencia del teorema 3.37, pues, para los primos que cumplen $p \equiv -1 \pmod 4$, tenemos que $d = (4, p-1) = 2$, luego la mitad de las clases de restos son restos bicuadráticos y, como todo resto bicuadrático es obviamente cuadrático, concluimos que los restos cuadráticos son los mismos que los bicuadráticos y la situación es trivial en este caso.

En cambio, si $p \equiv 1 \pmod 4$, entonces $d = (4, p-1) = 4$, por lo que el teorema 3.37 nos da que sólo la cuarta parte de los restos módulo p son restos bicuadráticos, y para investigarlos vamos a definir un símbolo bicuadrático que asignará a cada entero, no una de las dos potencias de -1, como en el caso cuadrático, ni una de las tres potencias de ζ, como en el caso cúbico, sino una de las cuatro potencias de i, de modo que la reciprocidad bicuadrática está

estrechamente relacionada con la aritmética del anillo $\mathbb{Z}[i]$ de los enteros de Gauss. En realidad en este capítulo no vamos a introducir ideas nuevas, sino que meramente adaptaremos las que hemos empleado en el capítulo anterior para tratar con la ley de reciprocidad cúbica.

16.1 El símbolo potencial bicuadrático

Vamos a probar los resultados previos y a definir los conceptos necesarios para enunciar la ley de reciprocidad bicuadrática, para lo cual seguimos de cerca la estructura del capítulo precedente. Como ya hemos señalado, tenemos que sustituir los enteros de Eisenstein por los de Gauss. Recordemos que el anillo $\mathbb{Z}[i]$ tiene cuatro unidades, $\pm 1, \pm i$, por lo que cada entero de Gauss tiene cuatro asociados. El único primo racional ramificado es $2 = -i\lambda^2$, donde $\lambda = 1 + i$.

Enteros primarios Vamos a definir enteros primarios que nos permitan seleccionar un asociado de cada entero de Gauss primo con λ. No podemos considerar congruencias módulo 2 porque sólo determinan 4 clases de restos, y sólo dos de ellas corresponden a las cuatro unidades. Necesitamos considerar congruencias módulo $\lambda^3 = 2i\lambda = 2i - 2$. Así, $2i \equiv 2 \pmod{\lambda^3}$ y $4 = -\lambda^4 \equiv 0 \pmod{\lambda^3}$, por lo que todo entero de Gauss $a + bi$ puede reducirse módulo λ^3 de modo que $0 \le a \le 3$ y $0 \le b \le 1$. Esto hace que las ocho clases de restos sean:

0	1	2	3
i	$1+i$	$2+i$	$3+i$

o equivalentemente:

0	1	λ^2	-1
i	λ	$-i$	$i\lambda$

Así vemos que todo entero de Gauss primo con λ tiene que ser congruente módulo λ^3 con una de las cuatro unidades, por lo que multiplicándolo por la inversa de dicha unidad obtenemos un (único) asociado congruente con 1 módulo λ^3.

Definición 16.1 Diremos que un entero de Gauss α es *primario* si cumple $\alpha \equiv 1 \pmod{\lambda^3}$.

Acabamos de probar que todo entero de Gauss primo con λ tiene un único asociado primario. Más explícitamente, $a + bi$ es primario si $\lambda^3 \mid a - 1 + bi$, es decir, si

$$\frac{a-1+bi}{2(-1+i)} = -\frac{a-b-1}{4} - \frac{a+b-1}{4} \in \mathbb{Z}[i]$$

o, equivalentemente, si $a + b \equiv a - b \equiv 1 \pmod 4$. Es fácil ver que esto equivale a que

$$a \equiv 1 \pmod 4 \text{ y } b \equiv 0 \pmod 4 \quad \text{o bien} \quad a \equiv -1 \pmod 4 \text{ y } b \equiv 2 \pmod 4.$$

Ejemplo Si $\alpha = 4 + 7i$, usando que $2i \equiv 2 \pmod{i} \pmod{\lambda^3}$, obtenemos que $\alpha \equiv 4 + 6 + i = 10 + i \pmod{\lambda^3}$, y usando ahora que $4 \equiv 0 \pmod{\lambda^3}$, llegamos a que $\alpha \equiv 2 + i \equiv -i \pmod{\lambda^3}$, luego $i\alpha \equiv 1 \pmod{\lambda^3}$ y así $i\alpha = -7 + 4i$ es el asociado primario de α. ∎

El teorema siguiente se prueba exactamente igual que 15.3:

Teorema 16.2 *Un entero de Gauss no nulo ni unitario es primario si y sólo si se descompone en producto de factores primos primarios.*

El símbolo potencial bicuadrático Dejamos al lector la adaptación de la prueba del teorema 15.5 al caso bicuadrático:

Teorema 16.3 *Sea π un primo de Gauss no asociado a λ. Entonces, un entero de Gauss α no divisible entre π es un resto bicuadrático módulo π si y sólo si*

$$\alpha^{(\mathrm{N}(\pi)-1)/4} \equiv 1 \pmod{\pi}.$$

A su vez, siguiendo el razonamiento posterior a 15.5, vemos que todo entero de Gauss α cumple que $\alpha^{\mathrm{N}(\pi)-1} \equiv 1 \pmod{\pi}$, luego $\alpha^{(\mathrm{N}(\pi)-1)/4}$ es una de las raíces en el cuerpo $\mathbb{Z}[i]/(\pi)$ del polinomio $x^4 - 1$, pero dichas raíces son las clases de las unidades $\pm 1, \pm i$, luego:

Teorema 16.4 *Si π es un primo de Gauss que no divida a 2 y α es un entero de Gauss no divisible entre π, entonces existe un n tal que*

$$\alpha^{(\mathrm{N}(\pi)-1)/4} \equiv i^n \pmod{\pi}.$$

Esto nos permite definir el símbolo bicuadrático:

Definición 16.5 Si π es un primo de Gauss que no divida a 2 y α es un entero de Gauss arbitrario, definimos

$$\left(\frac{\alpha}{\pi}\right)_4 = \begin{cases} i^n & \text{si } \pi \nmid \alpha \text{ y } \alpha^{(\mathrm{N}(\pi)-1)/4} \equiv i^n \pmod{\pi}, \\ 0 & \text{si } \pi \mid \alpha. \end{cases}$$

Y es inmediato que este símbolo bicuadrático cumple las mismas propiedades básicas que el símbolo cúbico:

$$\alpha^{(\mathrm{N}(\pi)-1)/4} \equiv \left(\frac{\alpha}{\pi}\right)_4 \pmod{\pi},$$

$$\left(\frac{\alpha\beta}{\pi}\right)_4 = \left(\frac{\alpha}{\pi}\right)_4 \left(\frac{\beta}{\pi}\right)_4,$$

$$\overline{\left(\frac{\alpha}{\pi}\right)_4} = \left(\frac{\bar{\alpha}}{\bar{\pi}}\right)_4,$$

así como que $(\alpha/\pi)_4$ sólo depende del resto de α módulo π. El teorema 16.3 se expresa ahora así:

Teorema 16.6 *Si π es un primo de Gauss que no divida a 2 y α es un entero de Gauss no divisible entre π, entonces α es un resto bicuadrático módulo π si y sólo si $(\alpha/\pi)_4 = 1$.*

Ejemplo Consideremos $\pi = 3+2i$, que cumple $\mathrm{N}(\pi) = 13$ y $2i \equiv 10 \pmod{\pi}$, luego $i \equiv 5 \pmod{\pi}$. Entonces

α	1	2	3	4	5	6	7	8	9	10	11	12
$\alpha^{(13-1)/4}$	1	8	1	12	8	8	5	5	1	12	5	12
$(\alpha/\pi)_4$	1	$-i$	1	-1	$-i$	$-i$	i	i	1	-1	i	-1

■

Nota Si $p = \mathrm{N}(\pi)$, entonces $\mathbb{Z}[i]/(\pi) = \mathbb{Z}_p$, por lo que un entero a es un resto bicuadrático módulo p (en $\mathbb{Z}$) si y sólo si lo es módulo π (en $\mathbb{Z}[i]$), si y sólo si $(a/\pi)_4 = 1$.

Sin embargo, para primos $p \equiv -1 \pmod 4$ hemos de tener cuidado, pues, por ejemplo, 3 no es un resto cuadrático (luego tampoco bicuadrático) módulo 7 (en $\mathbb{Z}$), pero $(3/7)_4 = 1$, ya que en $\mathbb{Z}[i]$ se cumple que $(1+i)^4 = -4 \equiv 3 \pmod 7$, luego 3 sí que es un resto bicuadrático módulo 7 en $\mathbb{Z}[i]$.

Pero en este caso ya hemos visto que no necesitamos considerar el símbolo bicuadrático, sino que un entero a es un resto bicuadrático módulo p si y sólo si es un resto cuadrático módulo p. ■

En general:

Teorema 16.7 *Dado un primo $p \equiv -1 \pmod 4$ y un entero a no divisible entre p, se cumple que*

$$\left(\frac{a}{p}\right)_4 = 1.$$

Demostración: Por la definición del símbolo potencial:

$$\left(\frac{a}{p}\right)_4 \equiv a^{(p^2-1)/4} = (a^{p-1})^{(p+1)/4} \equiv 1 \pmod q,$$

por el teorema de Fermat. ■

Definición 16.8 Sean α, β son enteros de Gauss, donde β no es nulo ni unitario y sea $\beta = \pi_1 \cdots \pi_r$ la descomposición de β es factores primos. Si ninguno de ellos divide a 2, definimos

$$\left(\frac{\alpha}{\beta}\right)_4 = \prod_{j=1}^{r} \left(\frac{\alpha}{\pi_j}\right)_4.$$

Si β es una unidad definimos $(\alpha/\beta)_4 = 1$.

Claramente el símbolo de Jacobi es multiplicativo en sus dos argumentos y depende únicamente del resto de α módulo β. Además,

$$\overline{\left(\frac{\alpha}{\beta}\right)_4} = \left(\frac{\bar{\alpha}}{\bar{\beta}}\right)_4.$$

Ahora podemos generalizar el teorema 16.7:

Teorema 16.9 *Si a, b son enteros no nulos primos entre sí y $b \neq \pm 1$ es impar, entonces*

$$\left(\frac{a}{b}\right)_4 = 1.$$

DEMOSTRACIÓN: Podemos suponer que $b > 0$. Sea $b = p_1 \cdots p_r q_1 \cdots q_s$ la descomposición en factores primos de b, donde $p_i \equiv 1$ (mód 4), $q_i \equiv -1$ (mód 4). Por 16.7 basta probar que $(a/p_i)_4 = 1$ o, equivalentemente, podemos suponer que $b \equiv 1$ (mód 4) es primo. Pongamos que $b = \pi\bar{\pi}$, de modo que

$$\left(\frac{a}{b}\right)_4 = \left(\frac{a}{\pi}\right)_4 \left(\frac{a}{\bar{\pi}}\right)_4 = \left(\frac{a}{\pi}\right) \overline{\left(\frac{a}{\pi}\right)_4} = 1. \qquad \blacksquare$$

La ley de reciprocidad bicuadrática Ya podemos enunciar la ley de reciprocidad:

Teorema 16.10 (Ley de reciprocidad bicuadrática) *Dados dos enteros de Gauss primarios $\alpha = a + bi$, $\beta = c + di$, se cumple que*

$$\left(\frac{\alpha}{\beta}\right)_4 = (-1)^{bd/4} \left(\frac{\beta}{\alpha}\right)_4$$

o, equivalentemente,

$$\left(\frac{\alpha}{\beta}\right)_4 = (-1)^{(\mathrm{N}(\alpha)-1)(\mathrm{N}(\beta)-1)/16} \left(\frac{\beta}{\alpha}\right)_4.$$

Leyes suplementarias

$$\left(\frac{i}{\alpha}\right)_4 = i^{(1-a)/2}, \qquad \left(\frac{-1}{\alpha}\right)_4 = (-1)^{(1-a)/2},$$
$$\left(\frac{\lambda}{\alpha}\right)_4 = i^{(a-b-b^2-1)/4}.$$

En particular $(2/\alpha)_4 = i^{-b/2}$.

La equivalencia entre los dos enunciados de la ley de reciprocidad se sigue de la congruencia:

$$\frac{\mathrm{N}(\alpha)-1}{4} \equiv \frac{b}{2} \text{ (mód 2)}.$$

En efecto, esto equivale a

$$\frac{a^2-1}{4}+\left(\frac{b}{2}\right)^2 \equiv \frac{b}{2} \text{ (mód 2)},$$

pero a es impar, luego $a^2 \equiv 1$ (mód 8), luego $(a^2-1)/4 \equiv 0$ (mód 2) y la conclusión es inmediata.

La primera ley suplementaria es fácil de probar:

Supongamos en primer lugar que $\alpha = \pi$ es primo. Entonces

$$\left(\frac{i}{\pi}\right)_4 = i^{(\mathrm{N}(\pi)-1)/4} = i^{(a^2+b^2-1)/2} = i^{(1-a)/2}.$$

La última igualdad se comprueba distinguiendo dos casos, o bien $a = 4k+1$, $b = 4r$, o bien $a = 4k+1$, $b = 4r+2$.

Si α es un entero de Gauss (primario) arbitrario, podemos factorizarlo como $\alpha = \pi_1 \cdots \pi_r$, donde cada π_j es un primo de Gauss primario. Razonando inductivamente, basta probar que si la igualdad se cumple para $\alpha, \alpha' \equiv 1$ (mód λ^3), entonces se cumple para $\alpha\alpha'$.

Digamos que $\alpha = a+bi$, $\alpha' = a'+b'i$, con lo que $\alpha\alpha' = aa'-bb'+(ab'+ba')i$. Basta ver que

$$i^{(1-a)/2+(1-a')/2} = i^{(1-aa'+bb')/2},$$

y esto se comprueba fácilmente distinguiendo los cuatro casos posibles, según si $a \equiv \pm 1$ (mód 4) y $a' \equiv \pm 1$ (mód 4). A su vez,

$$\left(\frac{-1}{\alpha}\right) = \left(\frac{i}{\alpha}\right)^2 = (-1)^{(1-a)/2}.$$

Ahora demostramos la segunda ley suplementaria admitiendo la ley de reciprocidad. En primer lugar suponemos que $b = 0$, con lo que la fórmula se reduce a

$$\left(\frac{\lambda}{a}\right)_4 = i^{(a-1)/4}.$$

Basta demostrarla cuando $a = \pm p$ es un primo racional (con el signo adecuado para que sea primario en $\mathbb{Z}[i]$, es decir, para que $\pm p \equiv 1$ (mód 4)). Esto se debe a que si $a_1, a_2 \equiv 1$ (mód 4), entonces, claramente,

$$\frac{a_1-1}{4}+\frac{a_2-1}{4} \equiv \frac{a_1a_2-1}{4} \text{ (mód 4)}.$$

Si $a = p \equiv 1$ (mód 4) (con $p > 0$), entonces $p \equiv \pi\bar{\pi}$ en $\mathbb{Z}[i]$, luego, teniendo en cuenta que $\lambda = 1+i = i(1-i) = i\bar{\lambda}$,

$$\left(\frac{\lambda}{a}\right)_4 = \left(\frac{\lambda}{\pi}\right)_4\left(\frac{\lambda}{\bar{\pi}}\right)_4 = \left(\frac{\lambda}{\pi}\right)_4\left(\frac{i}{\bar{\pi}}\right)_4\left(\frac{\bar{\lambda}}{\bar{\pi}}\right)_4 = \left(\frac{i}{\bar{\pi}}\right)_4\left(\frac{\lambda}{\pi}\right)_4\overline{\left(\frac{\lambda}{\pi}\right)_4} = \left(\frac{i}{\bar{\pi}}\right)_4$$

y la primera ley complementaria ya demostrada nos da la conclusión.

Supongamos ahora que $a = -p \equiv 1 \pmod 4$. Entonces

$$\lambda^p = (1+i)^p \equiv 1 + i^p = 1 - i \pmod p,$$

y tomando c de modo que $2c \equiv 1 \pmod p$, tenemos que

$$(1+i)(1-i)c \equiv 1 \pmod p,$$

luego, multiplicando la congruencia precedente por $(1-i)c$ resulta

$$\lambda^{p-1} \equiv c(1-i)^2 = -c2i \equiv -i = i^{-1} \pmod p.$$

Por lo tanto,

$$\left(\frac{\lambda}{-p}\right)_4 \equiv \lambda^{(\mathrm{N}(-p)-1)/4} = \lambda^{(p^2-1)/4} = (\lambda^{p-1})^{(p+1)/4} \equiv i^{(-p-1)/4} \pmod p.$$

Esto termina la prueba en el caso $b = 0$. En general, si $\alpha = a + bi$ y llamamos $d = (a, b)$, donde tomamos en d el signo adecuado para que $d \equiv 1 \pmod 4$, entonces $\alpha = d(a' + b'i)$, donde $(a', b') = 1$. Basta probar la fórmula para $\alpha' = a' + b'i$ (que es primario, porque d lo es), pues entonces

$$\left(\frac{\lambda}{\alpha}\right)_4 = \left(\frac{\lambda}{d}\right)_4 \left(\frac{\lambda}{\alpha'}\right)_4 = i^{(d-1)/4+(a'-b'-b'^2-1)/4} = i^{(d(a'-b'-b'^2)-1)/4}$$

$$= i^{(da'-db'-d^2b'^2-1)/4} = i^{(a-b-b^2-1)/4}.$$

en la tercera igualdad hemos usado de nuevo la congruencia

$$\frac{a_1-1}{4} + \frac{a_2-1}{4} \equiv \frac{a_1a_2-1}{4} \pmod 4,$$

mientras que en la cuarta hemos usado que $d \equiv 1 \equiv d^2 \pmod 4$, por lo que $db'^2/4 \equiv d^2b'^2/4 \pmod 4$. Equivalentemente, podemos suponer que $(a, b) = 1$. Como α es primario, se cumple que $a^* = i^b a = (-1)^{b/2}a \equiv 1 \pmod 4$. Entonces, usando 16.9,

$$\overline{\left(\frac{i}{a^*}\right)_4} \left(\frac{bi}{a^*}\right)_4 = \left(\frac{-i}{a^*}\right)_4 \left(\frac{bi}{a^*}\right) = \left(\frac{b}{a^*}\right)_4 = 1.$$

A su vez,

$$\left(\frac{\lambda}{a+bi}\right)_4 = \overline{\left(\frac{i}{a^*}\right)_4} \left(\frac{bi}{a^*}\right)_4 \left(\frac{\lambda}{a+bi}\right)_4 = i^{(a^*-1)/2} \left(\frac{a+bi}{a^*}\right)_4 \left(\frac{\lambda}{a+bi}\right)_4$$

$$= i^{(a^*-1)/2} \left(\frac{a^*}{a+bi}\right)_4 \left(\frac{\lambda}{a+bi}\right)_4 = i^{(a^*-1)/2} \left(\frac{i^b a}{a+bi}\right)_4 \left(\frac{\lambda}{a+bi}\right)_4$$

$$= i^{(a^*-1)/2} \left(\frac{i}{a+bi}\right)_4^b \left(\frac{a\lambda}{a+bi}\right)_4 = i^{(a^*-1)/2+b(1-a)/2} \left(\frac{a+ai}{a+bi}\right)_4$$

$$= i^{(a^*-1+b(1-a))/2}\left(\frac{i(a-b)}{a+bi}\right)_4 = i^{(a^*-1+b(1-a))/2}i^{(1-a)/2}\left(\frac{a-b}{a+bi}\right)_4$$

$$= i^{(a^*-a+b(1-a))/2}\left(\frac{a-b}{a+bi}\right)_4.$$

Ahora observamos que si se cumple $a \equiv 1 \pmod 4$ y $b \equiv 0 \pmod 4$ entonces

$$(a^*-a)/2 = 0 \equiv b^2/4 \pmod 4, \qquad b(1-a)/2 \equiv 0 \equiv b^2/2 \pmod 4,$$

mientras que si $a \equiv -1 \pmod 4$ y $b \equiv 2 \pmod 4$ igualmente

$$(a^*-a)/2 = -2a/2 \equiv 1 \equiv b^2/4 \pmod 4, \qquad b(1-a)/2 \equiv 2 \equiv b^2/2 \pmod 4,$$

luego

$$\left(\frac{\lambda}{a+bi}\right)_4 = i^{3b^2/4}\left(\frac{a+bi}{a-b}\right)_4 = i^{-b^2/4}\left(\frac{b+bi}{a-b}\right)_4 = i^{-b^2/4}\left(\frac{1+i}{a-b}\right)_4$$

$$= i^{-b^2/4}i^{(a-b-1)/4} = i^{(a-b-b^2-1)/4},$$

donde hemos usado el caso $b = 0$ probado previamente.

Por último:

$$\left(\frac{2}{\alpha}\right)_4 = \left(\frac{i^3\lambda^2}{\alpha}\right)_4 = \left(\frac{i}{\alpha}\right)_4^3\left(\frac{\lambda}{\alpha}\right)_4^2 = i^{3(1-a)/2}i^{(a-b-b^2-1)/2}$$
$$= i^{(2-2a-b-b^2)/2} = i^{-b/2}.$$

Notemos que la última igualdad equivale a que

$$1-a \equiv b\,\frac{b}{2} \pmod 4,$$

y, en efecto, si $a \equiv 1 \pmod 4$, entonces $b \equiv 0 \pmod 4$ y ambos miembros son 0 módulo 4, mientras que si $a \equiv -1 \pmod 4$, entonces $b \equiv 2 \pmod 4$, luego $b/2 \equiv 1 \pmod 2$, luego $b/2 \equiv \pm 1 \pmod 4$ y la congruencia se reduce al hecho trivial $2 \equiv \pm 2 \pmod 4$. ∎

16.2 Aplicaciones de la reciprocidad bicuadrática

El uso más elemental de la ley de reciprocidad bicuadrática es determinar si un primo dado es o no un resto bicuadrático módulo otro primo. Recordemos que el problema es trivial para módulo congruentes con -1 módulo 4, pues entonces los restos bicuadráticos coinciden con los cuadráticos.

Ejemplo *Estudiar si* 29 *es un resto bicuadrático módulo* 41.

Un divisor primo de 41 en $\mathbb{Z}[i]$ es $5+4i$. Así pues, hemos de calcular

$$\begin{aligned}\left(\frac{29}{5+4i}\right) &= \left(\frac{-3-i}{5+4i}\right) = \left(\frac{\lambda}{5+4i}\right)\left(\frac{-2+i}{5+4i}\right)\\ &= \left(\frac{-1}{5+4i}\right)\left(\frac{i}{5+4i}\right)\left(\frac{-1-2i}{5+4i}\right)\\ &= -\left(\frac{5+4i}{-1-2i}\right) = -\left(\frac{-\lambda}{-1-2i}\right)\\ &= -\left(\frac{-1}{-1-2i}\right)\left(\frac{\lambda}{-1-2i}\right) = -i,\end{aligned}$$

donde hemos realizado las divisiones euclídeas

$$29 = (5+4i)(4-3i)+(-3-i), \quad 5+4i = (-1-2i)(3i)+(-1-i)$$

y hemos tenido que pasar de $-2+i$ a su asociado $-1-2i \equiv 1 \pmod{\lambda^3}$ antes de aplicar la ley de reciprocidad. Concluimos que 29 no es un resto bicuadrático módulo 41. ∎

Ahora podemos dar una caracterización sencilla de los primos de la forma $p = x^2 + 64y^2$ (véase el segundo ejemplo de la página 515). En principio, para que p sea de esta forma es necesario que $p \equiv 1 \pmod 4$. Si se cumple esta condición, entonces $p = a^2 + b^2$, donde podemos elegir los signos de a y b de modo que $\pi = a + bi$ sea primario. Tenemos entonces que 2 es un resto bicuadrático módulo p si y sólo si $(2/\pi)_4 = i^{-b/2} = 1$, lo cual equivale a que $8 \mid b$ y, por lo tanto, a que p sea de la forma $p = x^2 + 64y^2$. Así pues:

> *Un primo p es de la forma $p = x^2 + 64y^2$ si y sólo si $p \equiv 1 \pmod 4$ y 2 es un resto bicuadrático módulo p.*

De aquí obtenemos una caracterización de los primos para los que 2 es un resto bicuadrático:

> *Si p es un primo impar, entonces 2 es un resto bicuadrático módulo p si y sólo si $p \equiv -1 \pmod 8$ o bien $p \equiv 1 \pmod 8$ y $p = x^2 + 64y^2$.*

Vamos a probar que la equivalencia se cumple tanto si $p \equiv 1 \pmod 4$ como si $p \equiv -1 \pmod 4$. En el primer caso, acabamos de ver que 2 es un resto bicuadrático módulo p si y sólo si $p = x^2 + 64y^2$, y esto ya implica la condición $p \equiv 1 \pmod 8$.

Si $p \equiv -1 \pmod 4$, sabemos que 2 es un resto bicuadrático módulo p si y sólo si es un resto cuadrático, lo cual equivale a que $p \equiv \pm 1 \pmod 8$, pero, siendo $p \equiv -1 \pmod 8$, esto equivale a que $p \equiv -1 \pmod 8$. ∎

También tenemos una versión bicuadrática del teorema 15.11:

Teorema 16.11 *Sea p un primo impar de la forma $p = a^2 + b^2$, con b par. Elegimos el signo de a de modo que $a \equiv 1 \pmod 4$ si y sólo si $4 \mid b$. Sea $q > 2$ un primo distinto de p, sea $q^* = (-1)^{(q-1)/2} q$ y definimos*

$$n = \frac{1}{4}\left(q - \left(\frac{-1}{q}\right)\right),$$

$$g_n = \sum_{\substack{1\le j\le n\\ \text{impar}}} \binom{n}{j} (-1)^{(j-1)/2}\, a^{n-j} b^j.$$

$$g_n^* = \sum_{\substack{0\le j\le n\\ \text{par}}} \binom{n}{j} (-1)^{j/2}\, a^{n-j} b^j.$$

Entonces q^* es un resto bicuadrático módulo p si y sólo si $q \mid g_n$. Si $(-1/q) = 1$, la misma condición vale para $-q^*$. En caso contrario, $-q^*$ es un resto bicuadrático módulo p si y sólo si $q \mid g_n^*$.

Demostración: En las condiciones del enunciado, tenemos que $p = \mathrm{N}(\pi)$, con $\pi = a + bi \equiv 1 \pmod 4$ sin más que elegir adecuadamente el signo de a. Por otra parte, $q^* \equiv 1 \pmod 4$. Sea $\epsilon = \pm 1$.

Supongamos en primer lugar que $q \equiv 1 \pmod 4$. Entonces $q^* = q = \rho\bar{\rho}$, donde podemos suponer que ρ es primario. Se cumple que ϵq^* es un resto bicuadrático módulo p si y sólo si $(\epsilon q^*/\pi)_4 = (\epsilon/\pi)_4 (q^*/\pi)_4 = 1$, pero, por la ley de reciprocidad,

$$\left(\frac{q^*}{\pi}\right)_4 = \left(\frac{\pi}{q^*}\right)_4 = \left(\frac{\pi}{\rho}\right)\left(\frac{\pi}{\bar{\rho}}\right)_4$$

y, razonando exactamente igual que en la prueba de 15.11, concluimos que ϵq^* es un resto bicuadrático módulo p si y sólo si

$$\pi^n \equiv \left(\frac{\epsilon}{\pi}\right)_4 \bar{\pi}^n \pmod \rho.$$

Supongamos ahora que $q \equiv -1 \pmod 4$, con lo que, al igual que en el caso anterior, $(\epsilon q^*/\pi) = (\epsilon/\pi)(\pi/q^*)$ y, por definición del símbolo potencial, ϵq^* es un resto bicuadrático módulo p si y sólo si

$$\pi^{(q^2-1)/4} \equiv \left(\frac{\epsilon}{\pi}\right)_4 \pmod q.$$

Exactamente igual que en la prueba de 15.11 concluimos que $\pi^{q-1} \equiv \bar{\pi}/\pi$, y con ello llegamos a que la congruencia anterior equivale a

$$\pi^n \equiv \left(\frac{\epsilon}{\pi}\right)_4 \bar{\pi}^n \pmod \rho.$$

Si en ambos casos llamamos ρ a un divisor de q (es decir, $\rho = q$ en el segundo caso), tenemos que q^* es un resto bicuadrático módulo p si y sólo si

$$(a + bi)^n \equiv \left(\frac{\epsilon}{\pi}\right)_4 (a - bi)^n \pmod \rho.$$

Desarrollando las potencias queda:

$$\sum_{j=1}^{n}\binom{n}{j}i^j a^{n-j}b^j \equiv \left(\frac{\epsilon}{\pi}\right)_4 \sum_{j=1}^{n}(-1)^j\binom{n}{j}i^j a^{n-j}b^j \pmod{\rho}.$$

Si $(\epsilon/\pi)_4 = 1$, los términos pares se cancelan y la congruencia se reduce a

$$\sum_{\substack{1\le j\le n\\ \text{impar}}}\binom{n}{j}(-1)^{(j-1)/2}ia^{n-j}b^j \equiv -\sum_{\substack{1\le j\le n\\ \text{impar}}}\binom{n}{j}(-1)^{(j-1)/2}ia^{n-j}b^j \pmod{\rho},$$

y esto equivale claramente a $q \mid g_n$. Si, por el contrario, $(\epsilon/\pi)_4 = -1$, son los términos impares los que se cancelan en la congruencia, y obtenemos

$$\sum_{\substack{1\le j\le n\\ \text{par}}}\binom{n}{j}(-1)^{j/2}a^{n-j}b^j \equiv -\sum_{\substack{1\le j\le n\\ \text{par}}}\binom{n}{j}(-1)^{j/2}a^{n-j}b^j \pmod{\rho},$$

y esto equivale a $q \mid g_n^*$. ∎

Veamos algunos casos particulares:

1. 2 *es un resto bicuadrático módulo* p *si y sólo si* $8 \mid b$.
2. -2 *es un resto bicuadrático módulo* p *si y sólo si* $4 \mid a-1$ *y* $8 \mid b$, *o bien* $4 \mid a+1$ *y* $8 \mid b-4$.
3. -3 *es un resto bicuadrático módulo* p *si y sólo si* $3 \mid b$.
4. 3 *es un resto bicuadrático módulo* p *si y sólo si* $4 \mid a-1$ *y* $3 \mid b$ *o bien* $4 \mid a+1$ *y* $3 \mid a$.
5. 5 *es un resto bicuadrático módulo* p *si y sólo si* $5 \mid b$.
6. -5 *es un resto bicuadrático módulo* p *si y sólo si* $4 \mid a-1$ *y* $5 \mid b$ *o bien* $4 \mid a+1$ *y* $5 \mid a$
7. -7 *es un resto bicuadrático módulo* p *si y sólo si* $7 \mid ab$.
8. 7 *es un resto bicuadrático módulo* p *si y sólo si* $4 \mid a-1$ *y* $7 \mid ab$ *o bien* $4 \mid a+1$ *y* $7 \mid a^2-b^2$.
9. -11 *es un resto bicuadrático módulo* p *si y sólo si* $11 \mid b(5a+b)(5a-b)$.
10. 11 *es un resto bicuadrático módulo* p *si y sólo si*

$$4 \mid a-1 \quad y \quad 11 \mid b(5a+b)(5a-b)$$

o bien

$$4 \mid a+1 \quad y \quad 11 \mid a(a+5b)(a-5b).$$

16.3 La prueba de la ley de reciprocidad

Sea π un primo de Gauss primario de norma prima p y llamemos χ_π al carácter módulo p dado por

$$\chi_\pi(x) = \left(\frac{x}{\pi}\right)_4.$$

Necesitamos el análogo siguiente al teorema 15.19:

Teorema 16.12 $J(\chi_\pi, \chi_\pi) = -(-1)^{(p-1)/4}\pi$.

Demostración: Tenemos que $J(\chi_\pi, \chi_p)$ es un entero de Gauss de norma $\mathrm{N}(J(\chi_\pi, \chi_\pi)) = |J(\chi_\pi, \chi_\pi)|^2 = p$, por el teorema 15.15. El mismo argumento empleado en la prueba del teorema 15.19 nos da que $\pi \mid J(\chi_\pi, \chi_\pi)$:

$$J(\chi_\pi, \chi_\pi) = \sum_{t\in\mathbb{Z}_p} \chi_\pi(t)\chi_\pi(1-t) \equiv \sum_{t=0}^{p-1} t^{(p-1)/4}(1-t)^{(p-1)/4}$$

$$= \sum_{t=0}^{p-1} t^{(p-1)/4} \sum_{j=0}^{(p-1)/4} \binom{(p-1)/4}{j}(-t)^j$$

$$= \sum_{j=0}^{(p-1)/4} \binom{(p-1)/4}{j}(-1)^j \sum_{t=0}^{p-1} t^{(p-1)/4+j} \pmod{\pi},$$

pero $(p-1)/4 + j < p-1$, luego si u es una raíz primitiva módulo p, se cumple que $u^{(p-1)/4+j} \neq 1$, y la última suma queda invariante módulo p cuando se multiplica por esta clase de restos, luego tiene que ser nula módulo p y, por consiguiente, $J(\chi_\pi, \chi_\pi) \equiv 0 \pmod{\pi}$.

Así, $J(\chi_\pi, \chi_\pi)$ es un asociado de π. Basta ver que $-(-1)^{(p-1)/4}J(\chi_\pi, \chi_\pi)$ es primario, pues, al ser asociado a π, que también lo es, tendrá que ser π.

Tenemos que

$$J(\chi_\pi, \chi_\pi) = \sum_{t\in\mathbb{Z}_p} \chi_\pi(t)\chi_\pi(1-t).$$

Observemos la situación cuando $p = 13$:

t	0	1	2	3	4	5	6	7	8	9	10	11	12
$1-t$	1	0	12	11	10	9	8	7	6	5	4	3	2

Vemos que los términos $\chi_\pi(t)\chi_\pi(1-t)$ se anulan para $t = 0, 1$, cada uno de ellos para $t = 2, \ldots, 6$ coincide con otro para $t = 8, \ldots, 12$ y el correspondiente a $t = 7$ aparece sólo una vez. Esto vale claramente para todo p, con lo que

$$J(\chi_\pi, \chi_\pi) = 2\sum_{t=2}^{(p-1)/2} \chi_\pi(t)\chi_\pi(1-t) + \chi_\pi((p+1)/2)^2.$$

Ahora observamos que todo entero de Gauss es congruente con 0 o con 1 módulo $\lambda = 1 + i$, y las unidades tienen que ser congruentes con 1, luego

$$\chi_\pi(t)\chi_\pi(1-t) \equiv 1 \pmod{\lambda},$$

luego

$$\sum_{t=2}^{(p-1)/2} \chi_\pi(t)\chi_\pi(1-t) \equiv \frac{p-3}{2} \text{ (mód } \lambda),$$

luego

$$2\sum_{t=2}^{(p-1)/2} \chi_\pi(t)\chi_\pi(1-t) \equiv p-3 \equiv -2 \text{ (mód } 2\lambda),$$

pues $2\lambda \mid 4 \mid p-1$, luego $p \equiv 1$ (mód 2λ). Por último, si $2c \equiv 1$ (mód p),

$$\chi_\pi((p+1)/2)^2 = \chi_\pi(c(p+1))^2 = \chi_\pi(c)^2 = \chi_\pi(2)^{-2} = \chi_\pi(-i\lambda^2)^2$$

$$= \chi_\pi((-i)^2)\chi_\pi(\lambda)^4 = \chi_\pi(-1) = (-1)^{(p-1)/4},$$

donde la última igualdad se sigue de la definición del símbolo bicuadrático. Uniendo todo esto obtenemos que

$$J(\chi_\pi, \chi_\pi) \equiv -2 + (-1)^{(p-1)/4} \text{ (mód } 2\lambda).$$

Por lo tanto

$$-(-1)^{(p-1)/4} J(\chi_\pi, \chi_\pi) \equiv 2(-1)^{(p-1)/4} - 1 \equiv 1 \text{ (mód } \lambda^3),$$

pues la última congruencia equivale a que $\pm 2 \equiv 2$ (mód λ^3), lo cual es cierto, pues $\lambda^3 \mid 4$. Con esto tenemos que $-(-1)^{(p-1)/4} J(\chi_\pi, \chi_\pi)$ es primario. ∎

De aquí obtenemos a su vez:

Teorema 16.13 $G(\chi_\pi)^4 = p\pi^2 = \pi^3\bar{\pi}$.

DEMOSTRACIÓN: Observemos que χ_π^2 es un carácter de orden 2 módulo p, luego tiene que ser el símbolo de Legendre

$$\chi_\pi^2(x) = \left(\frac{x}{p}\right).$$

Por el teorema 15.14, tenemos que

$$J(\chi_\pi, \chi_\pi) = \frac{G(\chi_\pi)^2}{G(\chi_\pi^2)}.$$

Por lo tanto:

$$J(\chi_\pi, \chi_\pi)^2 = \frac{G(\chi_\pi)^4}{G(\chi_\pi^2)^2} = \frac{G(\chi_\pi)^4}{p},$$

donde hemos usado (7.1), teniendo en cuenta que $p \equiv 1$ (mód 4). Ahora usamos el teorema 15.16:

$$J(\chi_\pi, \chi_\pi)^2 = \chi_\pi(-1)J(\chi_\pi, \chi_\pi)J(\chi_\pi, \chi_\pi^2),$$

luego

$$J(\chi_\pi, \chi_\pi) = \chi_\pi(-1)J(\chi_\pi, \chi_\pi^2).$$

Considerando de nuevo 15.16:

$$G(\chi_\pi)^4 = \chi(-1)pJ(\chi_\pi,\chi_\pi)J(\chi_\pi,\chi_\pi^2) = pJ(\chi_\pi,\chi_\pi)^2$$

y ahora basta aplicar el teorema anterior y tener en cuenta que $p = \pi\bar{\pi}$. ∎

Para probar la ley de reciprocidad bicuadrática consideramos en primer lugar el caso en que $\alpha = a$ es un entero racional, es decir, el caso $b = 0$. Entonces la fórmula se reduce a

$$\left(\frac{a}{\beta}\right)_4 = \left(\frac{\beta}{a}\right)_4.$$

Tenemos que $a \equiv 1 \pmod 4$, luego podemos descomponerlo en producto de primos $\pm q \equiv 1 \pmod 4$ y basta probar la fórmula para cada factor $\pm q$. (Descomponemos $(a/\beta)_4$ en producto de símbolos $(\pm q/\beta)_4$, aplicamos la ley de reciprocidad y volvemos a agruparlos.) Equivalentemente, podemos suponer que $\alpha = \pm q$, donde $q > 0$ es primo y $\pm q \equiv 1 \pmod 4$.

Similarmente, podemos descomponer β en producto de primos de Gauss primarios, y el mismo argumento nos permite suponer sin pérdida de generalidad que $\beta = \pi$ es un primo de Gauss primario. Si π también es un entero racional, entonces ambos miembros valen 1 por el teorema 16.9. Así pues, podemos suponer que $\mathrm{N}(\pi) = q \equiv 1 \pmod 4$. Llamemos $\chi = \chi_\pi$. El teorema 16.12 nos da que $J(\chi,\chi) = \pm\pi$.

Llamamos $\Omega = e^{2\pi i/4q}$, de modo que el anillo $\mathbb{Z}[\Omega]$ contiene a $\omega = e^{2\pi i/q} = \Omega^4$ y también a $i = \Omega^q$. Las congruencias módulo q que consideramos a continuación serán en este anillo.

Supongamos que $q \equiv -1 \pmod 4$. Entonces

$$J(\chi,\chi)^q \equiv \sum_{t\in\mathbb{Z}_p} \chi^q(t)\chi^q(1-t) \equiv \sum_{t\in\mathbb{Z}_p} \bar{\chi}(t)\bar{\chi}(1-t) = J(\bar{\chi},\bar{\chi}) \pmod q.$$

Por lo tanto,

$$\pi^{q+1} = J(\chi,\chi)^{q+1} \equiv J(\bar{\chi},\bar{\chi})J(\chi,\chi) = |J(\chi,\chi)|^2 = p \pmod q. \tag{16.1}$$

Por otro lado:

$$G^q(\chi) \equiv \sum_{t\in\mathbb{Z}_p} \chi^q(t)\omega^{qt} \equiv G_q(\chi^q) = \chi^{-q}(q)G(\chi^q) = \chi(q)G(\bar{\chi})$$

$$= \chi(-q)\overline{G(\chi)} \pmod q, \tag{16.2}$$

donde hemos usado 9.22, el hecho de que $q \equiv -1 \pmod 4$ y la igualdad (15.1), y usando este teorema:

$$G^{q+1}(\chi) \equiv \chi(-q)G(\chi)\overline{G(\chi)} = \chi(-q)p \pmod q.$$

Ahora usamos el teorema 16.13 y (16.1):

$$\chi(-q)\pi^{q+1} \equiv \chi(-q)p \equiv G^{q+1}(\chi) \equiv (G^4(\chi))^{(q+1)/4}$$

$$\equiv (p\pi^2)^{(q+1)/4} \equiv (\pi^{q+3})^{(q+1)/4} \pmod q.$$

En principio hemos probado esta congruencia en $\mathbb{Z}[\Omega]$, pero entonces tenemos que

$$\frac{\chi(-q)\pi^{q+1} - (\pi^{q+3})^{(q+1)/4}}{q} \in \mathbb{Z}[\Omega],$$

luego el cociente es a la vez un entero algebraico y un elemento de $\mathbb{Q}(i)$, luego está de hecho en $\mathbb{Z}[i]$, por lo que la congruencia anterior se cumple en $\mathbb{Z}[i]$.s Equivalentemente,

$$\left(\frac{-q}{\pi}\right)_4 = \chi(-q) \equiv \pi^{(q+3)(q+1)/4-(q+1)} = \pi^{(q^2-1)/4} \equiv \left(\frac{\pi}{-q}\right)_4 \text{ (mód } q).$$

Esto prueba la ley de reciprocidad cuando $q \equiv -1$ (mód 4). Ahora suponemos que $q \equiv 1$ (mód 4). La relación (16.2) es ahora:

$$G^q(\chi) \equiv \bar{\chi}(q)G(\chi) \text{ (mód } q),$$

luego

$$G^{q-1}(\chi) \equiv \bar{\chi}(q) \text{ (mód } q).$$

(Notemos que $G(\chi)$ tiene inverso módulo q, pues se cumple que $G(\chi)\overline{G(\chi)} = p$, y si $cp \equiv 1$ (mód q), entonces $c\overline{G(\chi)}$ es el inverso de $G(\chi)$.) Por 16.13:

$$\bar{\chi}(q) \equiv (G^4(\chi))^{(q-1)/4} \equiv (\pi^3\bar{\pi})^{(q-1)/4} \text{ (mód } q).$$

De nuevo, esta congruencia, probada en $\mathbb{Z}[\Omega]$, es válida de hecho en $\mathbb{Z}[i]$. Pongamos que $q = \rho\bar{\rho}$ en $\mathbb{Z}[i]$. Así la congruencia anterior implica que:

$$\overline{\left(\frac{q}{\pi}\right)_4} \equiv \left(\frac{\pi}{\rho}\right)_4^3 \left(\frac{\bar{\pi}}{\rho}\right)_4 = \overline{\left(\frac{\pi}{\rho}\right)_4} \left(\frac{\bar{\pi}}{\rho}\right)_4 = \text{ (mód } q).$$

Conjugando:

$$\left(\frac{q}{\pi}\right)_4 \equiv \left(\frac{\pi}{\rho}\right)_4 \left(\frac{\pi}{\bar{\rho}}\right)_4 = \left(\frac{\pi}{q}\right)_4.$$

Esto termina la prueba para el caso $b = 0$. En el caso general, ahora podemos suponer que $(a, b) = (c, d) = 1$. En efecto, en principio podemos descomponer $\alpha = d(a + bi)$, $\beta = e(c + di)$, donde $d \equiv e \equiv 1$ (mód 4) y $(a, b) = (c, d) = 1$. Entonces, admitiendo la ley de reciprocidad para $\alpha' = a + bi$, $\beta' = c + di$, tenemos que

$$\left(\frac{\alpha}{\beta}\right)_4 = \left(\frac{d}{e}\right)_4 \left(\frac{d}{\beta'}\right)_4 \left(\frac{\alpha'}{e}\right)_4 \left(\frac{\alpha'}{\beta'}\right)_4$$

$$= \left(\frac{e}{d}\right)_4 \left(\frac{\beta'}{d}\right)_4 \left(\frac{e}{\alpha'}\right)_4 \left(\frac{\beta'}{\alpha'}\right) (-1)^{bd/4} = \left(\frac{\beta}{\alpha}\right)_4 (-1)^{(db)(ed)/4}.$$

Para cada entero n impar, conviene definir $\sigma_n = (-1)^{(n-1)/2}$. Así $\sigma_a a$, $\sigma_c c$ y $\sigma_a\sigma_c(ac + bd)$ son primarios. Notemos además que

$$\sigma_c = (-1)^{(c-1)/2} = (-1)^{d/2} = i^d, \qquad \sigma_a = i^b.$$

Y también que $c \equiv -di$ (mód β), luego $c\alpha = ca + cbi \equiv ca + bd$ (mód β).

Por consiguiente:

$$\left(\frac{\sigma_c c}{\beta}\right)_4 \left(\frac{\alpha}{\beta}\right)_4 = \left(\frac{\sigma_c(ac+bd)}{\beta}\right)_4 = \left(\frac{\sigma_a}{\beta}\right)_4 \left(\frac{\sigma_a\sigma_c(ac+bd)}{\beta}\right)_4, \qquad (16.3)$$

donde hemos usado que, puesto que b es par,

$$\left(\frac{\sigma_a}{\beta}\right)_4^2 = \left(\frac{i}{\beta}\right)_4^{2b} = 1.$$

Por otro lado, por el caso $b = 0$ ya probado de la ley de reciprocidad y teniendo en cuenta 16.9:

$$\left(\frac{\sigma_c c}{\beta}\right)_4 = \left(\frac{\beta}{\sigma_c c}\right)_4 = \left(\frac{c+di}{\sigma_c c}\right)_4 = \left(\frac{di}{\sigma_c c}\right)_4 = \left(\frac{i}{\sigma_c c}\right)_4 = i^{(1-\sigma_c c)/2},$$

donde hemos usado la primera ley suplementaria, ya demostrada. Así (16.3) se convierte en

$$\left(\frac{\alpha}{\beta}\right)_4 = \left(\frac{i}{\beta}\right)_4^b \left(\frac{\beta}{\sigma_a\sigma_c(ac+bd)}\right)_4 i^{(\sigma_c c-1)/2}.$$

Del mismo modo se llega a:

$$\left(\frac{\beta}{\alpha}\right)_4 = \left(\frac{i}{\alpha}\right)_4^d \left(\frac{\alpha}{\sigma_a\sigma_c(ac+bd)}\right)_4 i^{(\sigma_a a-1)/2}.$$

Conjugamos, teniendo en cuenta que el primer símbolo es ± 1 porque d es par:

$$\overline{\left(\frac{\beta}{\alpha}\right)_4} = \left(\frac{i}{\alpha}\right)_4^d \left(\frac{\bar\alpha}{\sigma_a\sigma_c(ac+bd)}\right)_4 i^{(1-\sigma_a a)/2}.$$

Así:

$$\left(\frac{\alpha}{\beta}\right)_4 \overline{\left(\frac{\beta}{\alpha}\right)_4} = \left(\frac{i}{\beta}\right)_4^b \left(\frac{i}{\alpha}\right)_4^d \left(\frac{\beta\bar\alpha}{\sigma_a\sigma_c(ac+bd)}\right)_4 i^{(\sigma_c c-\sigma_a a)/2}.$$

Pero

$$\left(\frac{i}{\beta}\right)_4^b \left(\frac{i}{\alpha}\right)_4^d = i^{(1-c)b/2}\, i^{(1-a)d/2} = i^{bd/2} i^{bd/2} = 1,$$

pues, como α y β son primarios, $a+b \equiv c+d \equiv 1$ (mód 4). Por otra parte:

$$\left(\frac{\beta\bar\alpha}{\sigma_a\sigma_c(ac+bd)}\right)_4 = \left(\frac{ac+bd+(ad-bc)i}{\sigma_a\sigma_c(ac+bd)}\right)_4 = \left(\frac{(ad-bc)i}{\sigma_a\sigma_c(ac+bd)}\right)_4$$

El hecho de que α y β sean primos entre sí implica que todos los "numeradores" y "denominadores" de los símbolos que estamos considerando son primos entre sí (o de lo contrario los símbolos serían nulos). Por lo tanto, 16.9 nos da que

$$\left(\frac{\alpha}{\beta}\right)_4 \overline{\left(\frac{\beta}{\alpha}\right)_4} = \left(\frac{i}{\sigma_a\sigma_c(ac+bd)}\right)_4 i^{(\sigma_c c-\sigma_a a)/2}$$

$$= i^{(1-\sigma_a\sigma_c(ac+bd))/2} i^{(\sigma_c c-\sigma_a a)/2} = i^{(1-\sigma_a a\sigma_c c-\sigma_a a+\sigma_c c)/2}\, i^{-\sigma_a\sigma_c bd/2},$$

y haciendo $\sigma_a a = 1 + 4k$, $\sigma_c c = 1 + 4l$ queda

$$1 - (1 + 4k)(1 + 4l) - 1 - 4k + 1 + 4l = -8k - 16kl,$$

luego la primera potencia de i vale 1 y llegamos a que

$$\left(\frac{\alpha}{\beta}\right)_4 \overline{\left(\frac{\beta}{\alpha}\right)_4} = (-1)^{-\sigma_a \sigma_c bd/4} = (-1)^{bd/4}.$$

■

Capítulo XVII

Enteros ciclotómicos

En el capítulo VIII señalábamos que, sin la base algebraica adecuada, no estamos en condiciones de estudiar anillos de enteros algebraicos más allá de los correspondientes a cuerpos cuadráticos. No obstante, apurando las técnicas de las que disponemos, aquí estudiaremos una familia de anillos de enteros algebraicos de grado arbitrariamente grande: los enteros ciclotómicos de orden primo.

Con los resultados que obtendremos veremos varias aplicaciones. La primera será la demostración del Último Teorema de Fermat para exponente $p = 5$.

Euler demostró el caso $p = 3$ del Último Teorema de Fermat explotando la factorización:

$$z^3 = x^3 + y^3 = (x + y)(x^2 - xy + y^2).$$

Para $p = 5$ podemos factorizar análogamente

$$z^5 = x^5 + y^5 = (x + y)(x^4 - x^3y + x^2y^2 - xy^3 + y^4),$$

pero el segundo factor se vuelve poco menos que intratable. Dirichlet presentó una prueba de este caso en la que, mediante un cambio de variables y algunas manipulaciones, lograba pasar a una expresión algo más simple, a saber:

$$z^5 = 2p(p^4 + 10p^2q^2 + 5q^4).$$

Sin embargo, a la hora de obtener sistemáticamente pruebas para exponentes mayores, los argumentos de este tipo se vuelven cada vez más intrincados y sin futuro. En su lugar, Lamé propuso una estrategia diferente que Gauss aprovechó para simplificar la prueba de Euler para $p = 3$ y que, según veremos, funciona también para $p = 5$. Más aún, abre una vía para demostrar el Último Teorema de Fermat para un gran número de exponentes, aunque la generalización del argumento pasa por algunas consideraciones nada triviales.

La idea de Lamé (para el caso $p = 5$) consiste en considerar el número complejo[1] $\omega = e^{2\pi i/5}$, cuyas potencias $1, \omega, \omega^2, \omega^3, \omega^4$ son las raíces quintas de

[1] Véase [ITAn 5.8] y la discusión posterior.

la unidad en $\mathbb{C}$, es decir, las raíces del polinomio $T^5 - 1$. Por lo tanto, éste factoriza como

$$T^5 - 1 = (T-1)(T-\omega)(T-\omega^2)(T-\omega^3)(T-\omega^4).$$

Ahora sustituimos $T = -x/y$ y multiplicamos por $-y^5$, con lo que obtenemos:

$$x^5 + y^5 = (x+y)(x+\omega y)(x+\omega^2 y)(x+\omega^3 y)(x+\omega^4 y).$$

Esta ecuación será nuestro punto de partida para llegar a una contradicción, en completa analogía con lo que hemos visto en el caso $p = 3$, pero necesitamos algo de trabajo para aprovecharla convenientemente.

Otra aplicación que vamos a ver de los enteros ciclotómicos está relacionada con el problema que hemos discutido en la introducción al capítulo XI de [ITAn] sobre el valor de

$$\frac{p-1}{2}! \pmod p,$$

para cada primo p. En [ITAn 11.9] hemos probado que si $p \equiv -1 \pmod 4$ entonces

$$\frac{p-1}{2}! \equiv (-1)^{(h+1)/2} \pmod p,$$

donde h es el número de clases del cuerpo $\mathbb{Q}\big(\sqrt{-p}\,\big)$, pero hemos dejado sin analizar el caso en que $p \equiv 1 \pmod 4$. Lo que hemos probado es que en tal caso el factorial de $(p-1)/2$ tiene que ser una de las dos raíces cuadradas de -1 módulo p. Reproducimos la tabla incluida en la introducción de dicho capítulo, en la que hemos eliminado los valores de R y N, que ahora son irrelevantes, y hemos añadido la unidad fundamental del cuerpo $\mathbb{Q}\big(\sqrt{p}\,\big)$:

p	5	13	17	29	37	41	53	61	73
$\frac{p-1}{2}!$	2	5	-4	12	-6	9	23	11	27
h	1	1	1	1	1	1	1	1	1
ϵ	$\frac{1+\sqrt{5}}{2}$	$\frac{3+\sqrt{13}}{2}$	$\frac{8+2\sqrt{17}}{2}$	$\frac{5+\sqrt{29}}{2}$	$\frac{12+2\sqrt{37}}{2}$	$\frac{64+10\sqrt{41}}{2}$	$\frac{7+\sqrt{53}}{2}$	$\frac{39+5\sqrt{61}}{2}$	$\frac{2\,136+250\sqrt{73}}{2}$

Aquí demostraremos un resultado de Chowla al que aludíamos en [ITAn], según el cual, si la unidad fundamental es $\epsilon = u + v\sqrt{p}$, entonces

$$\frac{p-1}{2}! \equiv (-1)^{(h+1)/2} u \pmod p \tag{17.1}$$

(donde u puede ser semientero, es decir, de la forma $a/2$, y entonces en la congruencia hay que entender que $a/2$ es el producto de a por el inverso de 2 módulo p).

De este modo tenemos determinada la raíz de -1 determinada por el factorial de $(p-1)/2$ en términos de dos características aritméticas fundamentales del cuerpo $\mathbb{Q}\big(\sqrt{p}\,\big)$: su número de clases y su unidad fundamental.

La prueba se basará en las fórmulas obtenidas analíticamente en [ITAn] para el número de clases de un cuerpo cuadrático y en una reinterpretación los cálculos que hicimos en la prueba del teorema 7.10 para determinar el signo de las sumas de Gauss cuadráticas en términos de enteros ciclotómicos.

17.1 Números ciclotómicos

Si $n \geq 2$ es un número natural, sabemos que en $\mathbb{C}$ hay exactamente n raíces de la unidad, que son los números $1, \omega, \omega^2, \ldots, \omega^{n-1}$, donde $\omega = e^{2\pi i/n}$. Estos números son algebraicos (de hecho enteros algebraicos), pues son raíces del polinomio $x^n - 1$. Podemos considerar entonces el polinomio mínimo de ω, definido en 8.8, es decir, el menor polinomio con coeficientes racionales que tiene por raíz a ω. Ciertamente, no es $x^n - 1$, pues este polinomio no es irreducible:

$$x^n - 1 = (x-1)(x^{n-1} + x^{n-2} + \cdots + x + 1).$$

Definición 17.1 Para cada número natural $n \geq 1$, se llama *polinomio ciclotómico de orden n* al polinomio mínimo $c_n(x)$ del entero algebraico $\omega = e^{2\pi i/n}$.

El teorema 3.34 afirma que si p es un número primo, entonces el polinomio

$$c_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$$

es irreducible en $\mathbb{Q}[x]$, luego es necesariamente el polinomio ciclotómico de orden p (ya que el polinomio mínimo de un número algebraico ω es el único polinomio mónico irreducible en $\mathbb{Q}[x]$ que tiene raíz ω). Para órdenes no primos esto ya no es cierto. Por ejemplo:

$$x^4 - 1 = (x-1)(x+1)(x^2+1)$$

y, de estos tres factores, el que tiene por raíz a $i = e^{2\pi i/4}$ es el último, luego $c_4(x) = x^2 + 1$.

El cuerpo $\mathbb{Q}(\omega)$ de los *números ciclotómicos*[2] *de orden n* es el cuerpo dado por el teorema 8.12, cuyos elementos se expresan de forma única como

$$a_m\omega^m + a_{m-1}\omega^{m-1} + \cdots + a_1\omega + a_0,$$

donde m es el grado del polinomio $c_n(x)$ y los a_i son números racionales.

Hemos dado la definición para un orden n arbitrario para señalar que ya hemos estudiado a fondo los números ciclotómicos de órdenes 3 y 4, pues en estos casos tenemos que $\omega = \zeta = e^{2\pi i/3}$ o bien $\omega = i$ y los polinomios ciclotómicos tienen ambos grado 2:

$$c_3(x) = x^2 + x + 1, \qquad c_4(x) = x^2 + 1,$$

por lo que los cuerpos de números ciclotómicos son los cuerpos de cocientes $\mathbb{Q}(\zeta)$ y $\mathbb{Q}(i)$ de los anillos de enteros de Eisenstein y de Gauss, respectivamente. El caso $n = 2$ es trivial, pues para este orden $\omega = -1$, $c_2(x) = x + 1$ y los números ciclotómicos son simplemente los números racionales.

Sin embargo, a partir de aquí vamos a trabajar exclusivamente con los cuerpos de números ciclotómicos de orden primo $p \geq 3$.

[2]"ciclotómico" significa "que divide al círculo", y su nombre se debe a que, geométricamente, las potencias de ω dividen a la circunferencia unidad en n partes iguales.

17.2 Enteros ciclotómicos de orden 5

Dado que hasta ahora sólo hemos trabajado sistemáticamente con cuerpos cuadráticos, empezaremos estudiando explícitamente el cuerpo de los números ciclotómicos de orden $p = 5$ y luego extenderemos los conceptos y resultados que obtendremos al caso general, en la medida de lo posible.

Vamos a prescindir incluso de las definiciones que hemos dado en la sección precedente, y el uso que hemos hecho de los teoremas 3.34 y 8.12, pues es fácil obtener los mismos hechos mediante argumentos directos.

Para ello en esta sección será siempre $\omega = e^{2\pi i/5}$ y definimos los *números ciclotómicos* (de orden 5) como los números complejos de la forma $f(\omega)$, donde $f(x)$ es un polinomio con coeficientes racionales. Los números de la forma $f(\omega)$, donde f tiene coeficientes enteros, se llaman *enteros ciclotómicos* (de orden 5). Representaremos por $\mathbb{Q}(\omega)$ al conjunto de todos los números ciclotómicos y $\mathbb{Z}[\omega]$ al conjunto de los enteros ciclotómicos. Pronto veremos que $\mathbb{Q}(\omega)$ así definido es el mismo definido en la sección anterior.

Ejemplo El número complejo

$$\alpha = \omega^9 + 2\omega^7 + 4\omega^5 + \omega^2 - 3$$

es un ejemplo de entero ciclotómico. Ahora bien, puesto que $\omega^5 = 1$, tenemos también que $\omega^7 = \omega^2$ y $\omega^9 = \omega^4$, luego

$$\alpha = \omega^9 + 2\omega^7 + 4\omega^5 + \omega^2 - 3 = \omega^4 + 2\omega^2 + 4 + \omega^2 - 3 = \omega^4 + 3\omega^2 + 1.$$

En general, aunque la definición de número ciclotómico permite que el polinomio $f(x)$ tenga grado arbitrario, en la práctica todo número ciclotómico puede expresarse en la forma $f(\omega)$, donde el polinomio $f(x)$ tiene grado menor o igual que 4, ya que cualquier potencia de ω de grado 5 o superior se puede reducir a otra de grado a lo sumo 4. ∎

Así pues, todo número (resp. entero) ciclotómico admite una expresión en la forma

$$a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0, \tag{17.2}$$

donde a_0, a_1, a_2, a_3, a_4 son números racionales (resp. enteros). Cuando lo expresemos de esta forma diremos que está dado en *forma canónica.*

Hemos definido los números ciclotómicos admitiendo polinomios de grado arbitrario porque así es inmediato que la suma y el producto de números (resp. de enteros) ciclotómicos es de nuevo un número (resp. un entero) ciclotómico. Basta tener en cuenta que

$$f(\omega) + g(\omega) = (f + g)(\omega), \qquad f(\omega)g(\omega) = (fg)(\omega).$$

Más aún, es inmediato que $\mathbb{Q}(\omega)$ y $\mathbb{Z}[\omega]$ son subanillos del cuerpo $\mathbb{C}$ de los números complejos (sólo hay que comprobar además que 0 y 1 son enteros ciclotómicos y que el opuesto de un número/entero ciclotómico es también un número/entero ciclotómico).

Sin embargo, para operar con ellos podemos expresarlos siempre en forma canónica. Para sumar y multiplicar números ciclotómicos dados en forma canónica podemos usar adaptaciones obvias de los algoritmos que usamos para sumar y multiplicar números enteros. Por ejemplo:

$$\begin{array}{rrrrr}
2\omega^4 & +\,\omega^3 & & -\omega & +1\\
+\,\omega^4 & -2\omega^3 & +\omega^2 & +3\omega & +5\\
\hline
3\omega^4 & -\omega^3 & +\omega^2 & +2\omega & +6
\end{array}$$

$$\begin{array}{rrrrrrrrr}
 & & & & 2\omega^4 & +\omega^3 & & -\omega & +1\\
 & & & & \times\ \omega^4 & -2\omega^3 & +\omega^2 & +3\omega & +5\\
\hline
 & & & & 10\omega^4 & +5\omega^3 & & -5\omega & +5\\
 & & & 6\omega^5 & +3\omega^4 & & -3\omega^2 & +3\omega & \\
 & & 2\omega^6 & +\omega^5 & & -\omega^3 & +\omega^2 & & \\
 & -4\omega^7 & -2\omega^6 & & +2\omega^4 & -2\omega^3 & & & \\
+\ 2\omega^8 & +\omega^7 & & -\omega^5 & +\omega^4 & & & & \\
\hline
2\omega^8 & -3\omega^7 & & +6\omega^5 & +16\omega^4 & +2\omega^3 & -2\omega^2 & -2\omega & +5
\end{array}$$

El resultado no está en forma canónica, pero podemos escribir los términos en $\omega^5 = 1$ en la primera columna, los términos con $\omega^6 = \omega$ en la segunda columna, etc. Así:

$$\begin{array}{rrrrrr}
 & 2\omega^4 & +\omega^3 & & -\omega & +1\\
\times & \omega^4 & -2\omega^3 & +\omega^2 & +3\omega & +5\\
\hline
 & 10\omega^4 & +5\omega^3 & & -5\omega & +5\\
 & +3\omega^4 & & -3\omega^2 & +3\omega & +6\omega^5\\
 & & -\omega^3 & +\omega^2 & +2\omega^6 & +\omega^5\\
 & +2\omega^4 & -2\omega^3 & -4\omega^7 & -2\omega^6 & \\
+ & +\,\omega^4 & +2\omega^8 & +\omega^7 & & -\omega^5\\
\hline
 & 16\omega^4 & +4\omega^3 & -5\omega^2 & -2\omega & +11
\end{array}$$

con lo que obtenemos directamente el resultado en forma canónica. Finalmente, es más práctico trabajar únicamente con los coeficientes, omitiendo las potencias de ω, que se deducen de la posición de cada número:

$$\begin{array}{rrrrrr}
 & 2 & 1 & 0 & -1 & 1\\
\times & 1 & -2 & 1 & 3 & 5\\
\hline
 & 10 & 5 & 0 & -5 & 5\\
 & 3 & 0 & -3 & 3 & 6\\
 & 0 & -1 & 1 & 2 & 1\\
 & 2 & -2 & -4 & -2 & 0\\
+ & 1 & 2 & 1 & 0 & -1\\
\hline
 & 16 & 4 & -5 & -2 & 11
\end{array}$$

Ejemplo Consideremos ahora los números ciclotómicos

$$\omega^3 + 3\omega, \qquad 2\omega^4 + 3\omega^3 + 2\omega^2 + 5\omega + 2, \qquad -\frac{1}{2}\omega^4 + \frac{1}{2}\omega^3 - \frac{1}{2}\omega^2 + \frac{5}{2}\omega - \frac{1}{2}.$$

Vemos que los dos primeros son enteros, mientras que el tercero no lo es.

Si el lector está de acuerdo con esta última afirmación, deberá prestar atención a este ejemplo, pues en realidad sucede que los tres números son el mismo entero ciclotómico. En efecto, esto se debe a que

$$x^5 - 1 = (x-1)(x^4 + x^3 + x^2 + x + 1)$$

y, como ω es raíz del polinomio de la izquierda, pero no de $x - 1$, tiene que ser raíz del polinomio de la derecha, que no es sino el que en la sección anterior hemos definido como $c_5(x)$:

$$\omega^4 + \omega^3 + \omega^2 + \omega + 1 = 0.$$

Como consecuencia, si r es cualquier número racional, se cumple que

$$\begin{aligned} a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0 &= a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0 + r\cdot 0 \\ &= a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0 + r(\omega^4 + \omega^3 + \omega^2 + \omega + 1) \\ &= (a_4 + r)\omega^4 + (a_3 + r)\omega^3 + (a_2 + r)\omega^2 + (a_1 + r)\omega + a_0 + r. \end{aligned}$$

Vemos así que un número ciclotómico, expresado en forma canónica (17.2), permanece inalterado si sumamos un mismo número racional a todos sus coeficientes. En el ejemplo precedente, el "segundo" número ciclotómico es el que resulta de sumar 2 a todos los coeficientes del primero, por lo que no es otro, sino el mismo, y el "tercero" resulta de restar 1/2 a los coeficientes del primero, por lo que también se trata del mismo entero ciclotómico. ∎

En particular, restando a_4 a todos los coeficientes de un número ciclotómico expresado en forma canónica conseguimos una expresión en la que $a_4 = 0$. Cuando esto sucede diremos que el número está expresado en *forma reducida.*

La forma reducida se pierde al aplicar el algoritmo del producto, por lo que en la práctica es más cómodo operar en forma canónica y, si acaso, reducir el resultado final.

Vamos a probar que cada entero ciclotómico admite una única expresión en forma reducida, pero antes conviene considerar el ejemplo siguiente:

Ejemplo Definimos

$$\eta = \omega^4 + \omega, \qquad \eta' = \omega^3 + \omega^2.$$

Es fácil ver que

$$\eta + \eta' = \eta\eta' = \omega^4 + \omega^3 + \omega^2 + \omega = -1.$$

Por lo tanto,

$$(x - \eta)(x - \eta') = x^2 - (\eta + \eta')x + \eta\eta' = x^2 + x - 1,$$

pero las raíces de este polinomio son $(-1 \pm \sqrt{5})/2$.

Por otra parte, como $\omega \cdot \omega^4 = 1$, se cumple que $\omega^4 = \omega^{-1} = \bar{\omega}$, luego $\eta = \omega + \bar{\omega} = 2\,\mathrm{Re}\,\omega = 2\cos(2\pi/5) > 0$, por lo que

$$\eta = \frac{-1+\sqrt{5}}{2} = \omega^4 + \omega, \qquad \eta' = \omega^3 + \omega^2 = \frac{-1-\sqrt{5}}{2}.$$

Vemos así que $\sqrt{5} = 2\eta + 1 = 2\omega^4 + 2\omega + 1$ es un entero ciclotómico de orden 5, al igual que

$$\epsilon = \frac{1+\sqrt{5}}{2} = -\eta' = -\omega^3 - \omega^2.$$

Es claro entonces que el cuerpo cuadrático $\mathbb{Q}\big(\sqrt{5}\,\big)$ está contenido en el anillo $\mathbb{Q}(\omega)$ de los números ciclotómicos (de orden 5) y su anillo de enteros $\mathbb{Z}[\epsilon]$ está contenido en el anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos. ∎

La unicidad de la forma reducida de un número ciclotómico es consecuencia del teorema siguiente:

Teorema 17.2 *El número ω no es raíz de ningún polinomio no nulo de grado menor que 4 con coeficientes racionales.*

Demostración: Supongamos que $p(x)$ es un polinomio de grado ≤ 3 con coeficientes racionales tal que $p(\omega) = 0$. Ya hemos observado que $\omega^4 = \omega^{-1} = \bar{\omega}$, luego si $p(\omega) = 0$ también $p(\omega^4) = 0$, luego $p(x)$ tiene al menos dos raíces, luego tiene grado al menos 2. Si tuviera grado 2 tendría que ser

$$p(x) = a(x-\omega)(x-\omega^4) = a(x^2 - \eta x + 1),$$

pero entonces η sería racional, y no lo es. Por lo tanto, el grado tiene que ser 3. Dividimos:

$$x^4 + x^3 + x^2 + x + 1 = p(x)c(x) + r(x),$$

donde $r(x)$ es nulo o tiene grado ≤ 2. Al evaluar en ω vemos que $r(\omega) = 0$, pero hemos demostrado que ω no es raíz de polinomios no nulos con coeficientes racionales de grado ≤ 2, luego tiene que ser $r = 0$, de modo que

$$p(x) \mid x^4 + x^3 + x^2 + x + 1 = (x-\omega)(x-\omega^2)(x-\omega^3)(x-\omega^4),$$

pero entonces $p(x)$ tiene que tener como raíces ω, ω^4 y una más entre ω^2 y ω^3, pero esto es imposible, ya que $\omega^3\omega^2 = 1$, por lo que $\omega^3 = \bar{\omega}^2$ y así, si ω^2 es raíz de $p(x)$, lo mismo tiene que sucederle a su conjugado ω^3, y viceversa, luego $p(x)$ tendría que tener cuatro raíces y tenemos una contradicción. ∎

Esto nos da una demostración directa de que el polinomio ciclotómico $c_5(x)$ es irreducible en $\mathbb{Q}[x]$, ya que si pudiera descomponerse en producto de dos o más factores irreducibles, ω tendría que ser raíz de alguno de ellos, y hemos probado que esto es imposible.

Ahora podemos concluir que $c_5(x)$ no sólo tiene a ω entre sus raíces, sino que divide a cualquier otro polinomio $f(x) \in \mathbb{Q}[x]$ que tenga raíz ω.

Esto es lo que afirma el teorema 8.7, pero podemos probarlo directamente: consideramos el máximo común divisor de $c_5(x)$ y $f(x)$, que no puede ser más que 1 o $c_5(x)$, y basta probar que no puede ser 1, pero es que si fuera 1, por la relación de Bezout, existirían polinomios tales que $u(x)c_5(x) + v(x)f(x) = 1$, y evaluando en ω obtendríamos que $1 = 0$.

Como consecuencia:

Teorema 17.3 *Todo número ciclotómico se expresa de forma única en forma reducida, de modo que dos números ciclotómicos son iguales si y sólo si sus formas reducidas tienen los mismos coeficientes, si y sólo si sus coeficientes en forma canónica se diferencian en un sumando racional constante.*

DEMOSTRACIÓN: Si $f(\omega) = g(\omega)$, donde $f(x)$ y $g(x)$ son dos polinomios con coeficientes racionales de grado ≤ 4, tenemos que $(f - g)(x)$ es un polinomio de grado ≤ 4 que tiene a ω por raíz. Por el teorema anterior $f(x)-g(x) = c_5(x)q(x)$, para cierto polinomio $q(x)$, pero como $c_5(x)$ tiene grado 4 y $f(x) - g(x)$ tiene grado ≤ 4, el polinomio $q(x)$ tiene que ser constante, digamos $q(x) = r$, para cierto número racional r, de modo que

$$f(x) = g(x) + r(x^4 + x^3 + x^2 + x + 1),$$

luego los coeficientes de f y los de g se diferencian en una constante r. Si ambos están en forma canónica, entonces $r = 0$ (es la diferencia entre sus coeficientes de grado 4, que son ambos nulos), luego todos los coeficientes son iguales. ∎

Así pues, ahora es evidente que, por ejemplo,

$$4\omega^3 - 3\omega^2 + 1 \neq 2\omega^3 + 7,$$

puesto que ambos términos están en forma reducida.

Hemos visto un ejemplo según el cual un entero ciclotómico admite expresiones en forma canónica con coeficientes no enteros. Sin embargo, ahora podemos afirmar:

> *Un número ciclotómico es un entero ciclotómico si y sólo si su forma reducida tiene coeficientes enteros.*

En efecto, un entero ciclotómico es un número de la forma $f(\omega)$, donde $f(x)$ es un polinomio con coeficientes enteros. Reduciendo las potencias de ω exigir que f tenga grado menor o igual que 4 (también con coeficientes enteros), luego todo entero ciclotómico admite al menos una expresión en forma canónica con coeficientes enteros. Restando a_4 a dichos coeficientes obtenemos una forma reducida con coeficientes enteros, pero la forma reducida es única, luego obtenemos que la forma reducida de un entero ciclotómico tiene coeficientes enteros.

Otra consecuencia inmediata es:

> *Los únicos números racionales que son enteros ciclotómicos son los enteros racionales.*

En efecto, si r es un número racional, su expresión en forma reducida es simplemente r (es decir, con $a_0 = r$ y los demás coeficientes nulos), y acabamos de probar que si r es un entero ciclotómico, sus coeficientes en forma reducida tienen que ser enteros, luego r es entero.

He aquí otra consecuencia notable del teorema 17.2:

Teorema 17.4 *El anillo $\mathbb{Q}(\omega)$ de los enteros ciclotómicos (de orden 5) es un cuerpo.*

DEMOSTRACIÓN: Lo único que no es inmediato es que todo número ciclotómico tenga un inverso. Consideremos por ejemplo el número $\omega^2+2\omega+2$. Como número complejo que es, ciertamente tiene un inverso, a saber:

$$\frac{1}{\omega^2+2\omega+2}.$$

Lo que no es evidente es que este inverso sea también un número ciclotómico, y eso es lo que tenemos que demostrar. Lo probaremos en este caso en concreto, pero es obvio que el argumento es aplicable a cualquier número ciclotómico α no nulo.

Cualquier número ciclotómico no nulo dado puede expresarse en forma reducida, es decir como $f(\omega)$, donde f es un polinomio de grado a lo sumo 3. En nuestro ejemplo $f(x) = x^2+2x+2$. Como tiene grado menor que el polinomio ciclotómico $c_5(x)$ y éste es irreducible, ambos polinomios son primos entre sí, luego su máximo común divisor es 1 (o cualquier constante no nula). Aplicamos el algoritmo de Euclides para obtener polinomios que satisfagan la relación de Bezout:

c	r	u	v
	$x^4+x^3+x^2+x+1$	1	0
x^2-x+1	x^2+2x+2	0	1
$x+3$	$x-1$	1	$-x^2+x-1$
	5	$-x-3$	x^3+2x^2-2x+4

La conclusión es que

$$(-x-3)(x^4+x^3+x^2+x+1)+(x^3+2x^2-2x+4)(x^2+2x+2)=5.$$

Evaluando en ω queda que

$$(\omega^3+2\omega^2-2\omega+4)(\omega^2+2\omega+2)=5,$$

luego

$$\frac{1}{\omega^2+2\omega+2}=\frac{1}{5}\omega^3+\frac{2}{5}\omega^2-\frac{2}{5}\omega+\frac{4}{5}.$$

Este ejemplo muestra también que el inverso de un entero ciclotómico no tiene por qué ser entero. ∎

Con esto hemos demostrado directamente en el caso particular $p = 5$ todo lo que habíamos justificado para números ciclotómicos de cualquier orden primo p en la sección 17.1. Conviene señalar que hemos definido el anillo de los enteros ciclotómicos de orden 5 como el anillo $\mathbb{Z}[\omega]$ formado por los números ciclotómicos que en forma reducida tienen coeficientes enteros, eso sugiere que $\mathbb{Z}[\omega]$ es el anillo de enteros algebraicos del cuerpo $\mathbb{Q}(\omega)$, lo cual es cierto, pero todavía no lo hemos demostrado (ni nos hemos apoyado en ello en ningún momento).

17.3 Conjugaciones

Hemos visto que en el estudio de la aritmética de los enteros cuadráticos es fundamental el concepto de "norma". Enseguida veremos que es también posible definir la norma de un número ciclotómico, pero observemos en primer lugar que la generalización adecuada no es definir $\mathrm{N}(\alpha) = \alpha\bar{\alpha}$. Para empezar, habría que concretar a qué nos referimos cuando hablamos del conjugado $\bar{\alpha}$ de un número ciclotómico. Si pretendemos tomar como conjugación la conjugación compleja, es fácil ver que no funciona. Por ejemplo, en el caso $p = 5$, si $\alpha = \omega^2 + 2\omega + 2$, puesto que $\bar{\omega} = \omega^4$, tendríamos que

$$\begin{aligned}\mathrm{N}(\alpha) &= (\omega^2 + 2\omega + 2)(\omega^8 + 2\omega^4 + 2) = (\omega^2 + 2\omega + 2)(2\omega^4 + \omega^3 + 2)\\ &= -4\omega^3 - 4\omega^2 + 3.\end{aligned}$$

He aquí el cálculo:

$$\begin{array}{ccccc}
0 & 0 & 1 & 2 & 2\\
2 & 1 & 0 & 0 & 2\\
\hline
0 & 0 & 2 & 4 & 4\\
2 & 2 & 0 & 0 & 1\\
4 & 0 & 0 & 2 & 4\\
\hline
6 & 2 & 2 & 6 & 9
\end{array}$$

Vemos así que esta "norma" no es un entero racional, por lo que no es lo que necesitamos. Éste es el momento oportuno para reflexionar y profundizar en la noción algebraica de "conjugación".

Empezamos observando que la conjugación en el sentido de los números complejos no es la única conjugación que hemos venido manejando hasta ahora. Cuando consideramos cuerpos cuadráticos imaginarios, como $\mathbb{Q}\big(\sqrt{-3}\,\big)$, la conjugación que hemos considerado, es decir, $a+b\sqrt{-3} \mapsto a-b\sqrt{-3}$ coincide con la conjugación compleja si identificamos $\sqrt{-3} = \sqrt{3}i$, pero en los cuerpos cuadráticos reales, como $\mathbb{Q}\big(\sqrt{3}\,\big)$, hemos considerado la conjugación $a+b\sqrt{3} \mapsto a-b\sqrt{3}$, que no es la conjugación compleja.

¿Qué hace que "lo correcto" sea considerar que el conjugado de $\sqrt{-3}$ es $-\sqrt{-3}$, o que el conjugado de $\sqrt{3}$ es $-\sqrt{3}$, o que el conjugado de i es $-i$? Consideremos, por ejemplo, el último caso:

En el cuerpo $\mathbb{C}$ de los números complejos hay dos raíces cuadradas de -1. Si llamamos i a una de ellas, la otra es $-i$, pero no hay ningún criterio algebraico que determine cuál de las dos raíces debe ser llamada i y cuál $-i$.

Imaginemos dos matemáticos A y B, el primero de los cuales ha llamado i a una raíz cuadrada de -1, mientras que el segundo ha decidido trabajar con la otra raíz cuadrada, a la que llama j, de modo que $j = -i$. El primer matemático expresa los números complejos como $a + bi$, mientras que el segundo los llama $a + bj$. El primero los suma y los multiplica con las reglas

$$(a+bi)+(c+di) = a+c+(b+d)i, \qquad (a+bi)(c+di) = ac-bd+(ad+bc)i,$$

mientras que el segundo usa las reglas

$$(a+bj)+(c+dj) = a+c+(b+d)j, \qquad (a+bj)(c+dj) = ac-bd+(ad+bc)j.$$

En definitiva, ambos hacen lo mismo, aunque hayan elegido raíces $i \neq j$. El número que el matemático A llama $3+5i$ para el matemático B es $3-5j$. Imaginemos ahora que ambos matemáticos llaman i a su unidad imaginaria, aunque sea una diferente en cada caso. Entonces sus reglas de suma y multiplicación son indistinguibles, y el conjugado del número que el matemático A llama $a+bi$ es el número $a-bi$ al que se refiere el matemático B cuando escribe $a+bi$. En otras palabras, el conjugado de un número $a+bi$ es el número al que habríamos llamado $a+bi$ si en la elección arbitraria de la raíz de -1 que nombramos como i hubiéramos hecho la elección opuesta. El hecho de que de haber elegido la otra raíz no cambia en nada las reglas de la suma y el producto (los matemáticos A y B usan las mismas reglas a pesar de haber elegido raíces distintas) es lo que confiere a la conjugación todas sus propiedades.

Lo mismo se aplica, por ejemplo, a la conjugación en el cuerpo $\mathbb{Q}\big(\sqrt{3}\big)$. El conjugado de un número $a + b\sqrt{3}$ es el número $a - b\sqrt{3}$ al que habríamos llamado $a + b\sqrt{3}$ si en la elección arbitraria de la raíz cuadrada de 3 a la que hemos decidido llamar $\sqrt{3}$ hubiéramos hecho la elección opuesta.

Pensemos ahora en lo que sucede en el cuerpo ciclotómico quinto $\mathbb{Q}(\omega)$. En el cuerpo $\mathbb{C}$ de los números complejos hay 5 raíces quintas de la unidad, de las cuales una es diferente del resto (el 1) pues tiene orden 1, mientras que las cuatro restantes tienen todas orden 5 y son algebraicamente indistinguibles. Nosotros hemos elegido llamar $\omega = e^{2\pi i/5}$, pero ¿qué ocurriría si otro matemático decidiera trabajar con $\zeta = \omega^3$ en lugar de con ω? Se encontraría con que las potencias de ζ son

$$\zeta^0 = 1, \quad \zeta^1 = \omega^3, \quad \zeta^2 = \omega, \quad \zeta^3 = \omega^4, \quad \zeta^4 = \omega^2,$$

de modo que el número ciclotómico que nosotros escribimos como

$$a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0$$

él lo escribiría como

$$a_4\zeta^3 + a_3\zeta + a_2\zeta^4 + a_1\zeta^2 + a_0.$$

Supongamos ahora que este matemático llamara también ω a la raíz que ha elegido (nuestra ω^3). Entonces el número que nosotros escribimos como

$$a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0,$$

él lo escribiría como

$$a_4\omega^3 + a_3\omega + a_2\omega^4 + a_1\omega^2 + a_0.$$

Así, podemos decir que $a_4\omega^3 + a_3\omega + a_2\omega^4 + a_1\omega^2 + a_0$ es el número que habríamos llamado $a_4\omega^4+a_3\omega^3+a_2\omega^2+a_1\omega+a_0$ si hubiéramos decidido llamar ω a la raíz que llamamos ω^3. Y podemos decir que los números

$$a_4\omega^3 + a_3\omega + a_2\omega^4 + a_1\omega^2 + a_0, \qquad a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0$$

son conjugados en el mismo sentido algebraico profundo en que lo son $a + bi$ y $a - bi$. La clave es que si un matemático decide llamar ω a lo que nosotros llamamos ω^3 expresará los números ciclotómicos igual que nosotros y los operará con las mismas reglas con que los operamos nosotros, y no se notará en nada que estamos llamando con la misma expresión a dos números distintos. Y el hecho de que no se note en nada cambiar ω por ω^3 (o en general, por cualquier potencia $\omega^j \neq 1$) a la hora de operar con números ciclotómicos hace que los cambios consistentes en sustituir $\omega \mapsto \omega^j$ satisfagan propiedades análogas a las de la conjugación compleja, y merezcan el nombre de "conjugaciones ciclotómicas". Vamos a precisar esta idea.

En el caso $p = 5$ podemos razonar así:

Sabemos que un mismo número ciclotómico admite expresiones distintas de la forma $f(\omega)$ con $f(x) \in \mathbb{Q}[x]$. Sin embargo, si se cumple que $f(\omega) = g(\omega)$, necesariamente se cumple también que $f(\omega^i) = g(\omega^i)$, para $i = 1, 2, 3, 4$.

En efecto, tenemos que el polinomio $f(x) - g(x)$ tiene por raíz a ω, luego $c_5(x) \mid f(x) - g(x)$, y esto implica que $f(x) - g(x)$ se anula también en cada ω^j.

Por consiguiente, podemos definir aplicaciones $\sigma_j : \mathbb{Q}(\omega) \longrightarrow \mathbb{Q}(\omega)$ dadas por $\sigma_j(f(\omega)) = f(\omega^j)$. Acabamos de probar que, si $\alpha \in \mathbb{Q}(\omega)$, el número $\sigma_j(\alpha)$ no depende de la expresión $\alpha = f(\omega)$ a partir de la cual lo calculamos.

Por ejemplo:

$$\begin{aligned}
\sigma_1(\omega^2 + 2\omega + 2) &= \omega^2 + 2\omega + 2,\\
\sigma_2(\omega^2 + 2\omega + 2) &= \omega^4 + 2\omega^2 + 2 = -\omega^3 + \omega^2 - \omega + 1,\\
\sigma_3(\omega^2 + 2\omega + 2) &= \omega + 2\omega^3 + 2 = 2\omega^3 + \omega + 2,\\
\sigma_4(\omega^2 + 2\omega + 2) &= \omega^3 + 2\omega^4 + 2 = -\omega^3 - 2\omega^2 - 2\omega.
\end{aligned}$$

Vemos que la conjugación σ_1 deja invariantes a todos los números ciclotómicos (en virtud de ella podemos decir que todo número ciclotómico es conjugado de sí mismo), mientras que σ_4 transforma ω en $\omega^4 = \bar{\omega}$, por lo que es la conjugación compleja usual, pero, además de estas dos, tenemos σ_2 y σ_3, y lo esencial es que las cuatro conjugaciones están en pie de igualdad, sin que podamos decir que ninguna de ellas es "la auténtica". No hay una (ni dos, si contamos la trivial), sino cuatro conjugaciones ciclotómicas de orden 5. De acuerdo con las explicaciones precedentes, $\sigma_3(\omega^2 + 2\omega + 2)$ es el número al que estaríamos llamando $\omega^2 + 2\omega + 2$ si hubiéramos decidido llamar ω al número al que de hecho llamamos ω^3.

Observemos que trivialmente se cumple que

$$\sigma_j(\alpha+\beta)=\sigma_j(\alpha)+\sigma_j(\beta), \qquad \sigma_j(\alpha\beta)=\sigma_j(\alpha)\sigma_j(\beta).$$

Por ejemplo, para el producto tenemos que si $\alpha=f(\omega)$ y $\beta=g(\omega)$, entonces $\alpha\beta=h(\omega)$, donde $h(x)=f(x)g(x)$. Por lo tanto,

$$\sigma_j(\alpha\beta)=h(\omega^j)=f(\omega^j)g(\omega^j)=\sigma_j(\alpha)\sigma_j(\beta).$$

Para la suma se razona igualmente, con el polinomio $h(x)=f(x)+g(x)$.

En general, para números ciclotómicos de orden primo p, las conjugaciones vienen dadas por la segunda parte del teorema 8.12:

Definición 17.5 Si $p>2$ es un número primo y $p\nmid j$, definimos la *conjugación* $\sigma_j:\mathbb{Q}(\omega)\longrightarrow\mathbb{Q}(\omega)$ como la aplicación dada por el teorema 8.12, determinada por que $\sigma_j(\omega)=\omega^j$, $\sigma_j(a)=a$, para todo número racional a y

$$\sigma_j(\alpha+\beta)=\sigma_j(\alpha)+\sigma_j(\beta), \qquad \sigma_j(\alpha\beta)=\sigma_j(\alpha)\sigma_j(\beta).$$

Ejercicio: Probar que $\sigma_j(-\alpha)=-\sigma_j(\alpha)$ y si $\alpha\neq 0$, entonces $\sigma_j(\alpha^{-1})=\sigma_j(\alpha)^{-1}$.

Notemos que hemos definido σ_j para todo número entero que cumpla $p\nmid j$, pero eso no significa que tengamos infinitas conjugaciones, sino que, si se cumple $j\equiv j'\pmod p$, entonces $\omega^j=\omega^{j'}$, luego la unicidad de las conjugaciones implica que $\sigma_j=\sigma_{j'}$.

Así pues, σ_j depende únicamente del resto de j módulo p, por lo que hay en realidad $p-1$ conjugaciones, una por cada clase de U_p.

La ventaja de considerar los índices en las conjugaciones σ_j como elementos de U_p es que entonces podemos afirmar que $\sigma_i\sigma_j=\sigma_{ij}$, donde el miembro izquierdo es la aplicación que resulta de hacer actuar primero σ_i y luego σ_j. Esto es trivial: dado un número ciclotómico $\alpha=f(\omega)$, tenemos que $\sigma_i(\alpha)=f(\omega^i)$, luego $\sigma_j(\sigma_i(\alpha))=f((\omega^j)^i)=f(\omega^{ij})=\sigma_{ij}(\alpha)$.

Por ejemplo, la tabla siguiente muestra el efecto de aplicar sucesivamente dos conjugaciones a un mismo número ciclotómico quinto:

	σ_1	σ_2	σ_3	σ_4
σ_1	σ_1	σ_2	σ_3	σ_4
σ_2	σ_2	σ_4	σ_1	σ_3
σ_3	σ_3	σ_1	σ_4	σ_2
σ_4	σ_4	σ_3	σ_2	σ_1

Así, la tabla afirma que $\sigma_3(\sigma_4(\alpha))=\sigma_2(\alpha)$, pues al aplicar σ_4 sustituimos ω por ω^4, y al aplicar σ_3 sustituimos ω por ω^3, luego cada potencia ω^4 se sustituye por $\omega^{12}=\omega^2$, luego el efecto es el mismo que si hubiéramos sustituido ω por ω^2 directamente.

Observemos que $\sigma_{-1}(\omega^i) = \omega^{-i} = \bar{\omega}^i$, donde la barra denota la conjugación compleja. De aquí se sigue fácilmente que $\sigma_{-1}(\alpha) = \bar{\alpha}$, para todo número ciclotómico α, luego de las $p-1$ conjugaciones de $\mathbb{Q}(\omega)$, una de ellas, σ_{-1} es precisamente la conjugación compleja usual, pero ahora ya no es la única, sino que es una más entre otras muchas.

Conviene pensar que cada número ciclotómico tiene $p-1$ conjugados (contándolo a él mismo), pero esto hay que entenderlo bien. Por ejemplo, para $p = 5$, en el caso de $\alpha = 2\omega^3 + 2\omega^2$ tenemos que

$$\begin{aligned}\sigma_1(2\omega^3+2\omega^2) &= 2\omega^3+2\omega^2,\\ \sigma_2(2\omega^3+2\omega^2) &= 2\omega+2\omega^4,\\ \sigma_3(2\omega^3+2\omega^2) &= 2\omega^4+2\omega,\\ \sigma_4(2\omega^3+2\omega^2) &= 2\omega^2+2\omega^3.\end{aligned}$$

por lo que sus "cuatro" conjugados son en realidad dos. Similarmente, los cuatro conjugados de $\beta = 7$ son iguales entre sí.

Los enteros ciclotómicos reales Como primera aplicación de lo visto hasta ahora vamos a determinar los números ciclotómicos reales de orden 5. En general, para un orden primo arbitrario p, tenemos que un número ciclotómico α será real si y sólo si $\alpha = \bar{\alpha}$, donde $\bar{\alpha}$ es el conjugado complejo, que ya sabemos que es $\sigma_{-1}(\alpha)$. Vamos a desarrollar este hecho en el caso $p = 5$. Entonces la conjugación compleja es σ_4. Si

$$\alpha = a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0,$$

tenemos que

$$\sigma_4(\alpha) = a_4\omega + a_3\omega^2 + a_2\omega^3 + a_1\omega^4 + a_0,$$

luego $\sigma_4(\alpha) = \alpha$ equivale a que $a_1 = a_4$ y $a_2 = a_3$ (notemos que, como α y $\sigma_4(\alpha)$ tienen el mismo coeficiente a_0, para que las expresiones en forma canónica coincidan todos los coeficientes tienen que coincidir, ya que tienen que diferenciarse en un número racional que necesariamente es 0). Equivalentemente, los números ciclotómicos fijados por σ_4 (los números ciclotómicos reales) son los de la forma

$$\alpha = c(\omega^4+\omega) + b(\omega^3+\omega^2) + a = c\eta + b\eta' + a.$$

Expresados en forma reducida son los de la forma $\alpha = b\eta' + a$. Pero sabemos que

$$\eta' = -\epsilon = -\frac{1+\sqrt{5}}{2},$$

luego los números ciclotómicos reales se expresan también (de forma única) como $a + b\epsilon$, con $a, b \in \mathbb{Q}$, pero éstos son precisamente los elementos del cuerpo cuadrático $\mathbb{Q}(\sqrt{5})$. Más aún, los enteros ciclotómicos reales son los números de esta forma con $a, b \in \mathbb{Z}$, que forman precisamente el anillo de enteros $\mathbb{Z}[\epsilon]$ de dicho cuerpo cuadrático. En resumen:

Teorema 17.6 *Los enteros ciclotómicos reales de orden* 5 *forman el cuerpo cuadrático* $\mathbb{Q}(\sqrt{5})$*, mientras que los enteros ciclotómicos reales forman el anillo de enteros de este cuerpo, es decir,* $\mathbb{Z}[\epsilon]$*, donde* $\epsilon = (1+\sqrt{5})/2$*.* ∎

Hemos determinado los números ciclotómicos invariantes por σ_4 y, trivialmente, los invariantes por σ_1 son todos los números ciclotómicos. Veamos ahora cuáles son los invariantes por σ_2 y σ_3. Si aplicamos σ_2 a un número ciclotómico arbitrario

$$\alpha = a_4\omega^4 + a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0$$

obtenemos:

$$\sigma_2(\alpha) = a_4\omega^3 + a_3\omega + a_2\omega^4 + a_1\omega^2 + a_0,$$

luego se cumple $\sigma_2(\alpha) = \alpha$ si y sólo si $a_4 = a_2$, $a_3 = a_4$, $a_2 = a_1$, $a_1 = a_3$, es decir, si y sólo si $\alpha = b(\omega^4+\omega^3+\omega^2+\omega)+a = -b+a$, si y sólo si α es racional. Igualmente se comprueba que lo mismo vale para σ_3:

Ejercicio: Estudiar la forma de los números de orden 7 fijados por cada una de las seis conjugaciones. ∎

En el caso de los cuerpos cuadráticos, los números fijados por la conjugación (no trivial) son precisamente los números racionales. Este hecho se corresponde con el teorema siguiente:

Teorema 17.7 *Los números ciclotómicos fijados por todas las conjugaciones son los números racionales.*

Demostración: Sea $\alpha = a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0$ un número ciclotómico expresado en forma canónica y supongamos que es invariante por todas las conjugaciones. Entonces:

$$\sigma_j(\alpha) = a_{p-1}\omega^{(p-1)j} + \cdots + a_1\omega^j + a_0 = a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0.$$

Como en el caso $p = 5$, los coeficientes de dos expresiones canónicas de un mismo número ciclotómico se tienen que diferenciar en un mismo número racional, pero como en este caso a_0 es el mismo en ambas, tienen que ser iguales, luego $a_1 = a_j$, para $j = 1, \ldots, p-1$, luego

$$\alpha = a(\omega^{p-1} + \cdots + \omega) + a_0 = a_0 - a \in \mathbb{Q}.$$

Obviamente, todo número racional es invariante por todas las conjugaciones. ∎

Nota Observemos que en el caso $p = 5$ hemos visto que, para que un número ciclotómico sea racional, no hace falta que sea invariante por todas las conjugaciones, sino que basta con que lo sea por σ_2 o por σ_3. Esto tiene una interpretación simple: si g es una raíz primitiva módulo p, entonces toda clase $j \in U_p$ es de la forma g^n, lo que se traduce en que σ_j se puede calcular aplicando n veces la conjugación σ_g, luego si un número ciclotómico α es fijado por σ_g, es fijado por todas las conjugaciones, luego es un número racional. ∎

17.4 La norma

Ahora ya estamos en condiciones de definir razonablemente la norma de un número ciclotómico. En el caso de los cuerpos cuadráticos, la norma está definida como $\mathrm{N}(\alpha) = \alpha\bar{\alpha}$, lo cual hasta ahora lo interpretábamos como "el producto de α por su conjugado", pero, según las consideraciones de la sección anterior, es "más profundo" interpretarlo como "el producto de los dos conjugados de α" y, si lo vemos así, la generalización natural a los números ciclotómicos es la siguiente:

Definición 17.8 Definimos la *norma* de un número ciclotómico α de orden p como

$$\mathrm{N}(\alpha) = \prod_{j\in U_p} \sigma_j(\alpha),$$

es decir, como el producto de todos sus conjugados.

Notemos que nos referimos al producto de todos los conjugados con posibles repeticiones. Por ejemplo, si $\alpha \in \mathbb{Q}$, todos sus conjugados son iguales, y tenemos que $\mathrm{N}(\alpha) = \alpha^{p-1}$.

En el teorema siguiente llamamos $\mathbb{Z}[\omega]$ al anillo formado por los números ciclotómicos de la forma $f(\omega)$, donde $f(x) \in \mathbb{Z}[x]$. Realizando la división euclídea $f(x) = c_5(x)q(x) + r(x)$, con $r(x) \in \mathbb{Z}[x]$, vemos que $f(x)$ puede tomar se de grado menor que $p-1$, con lo que los elementos de $\mathbb{Z}[\omega]$ son los que, expresados en forma reducida, tienen coeficientes enteros.

En la sección siguiente veremos que $\mathbb{Z}[\omega]$ es precisamente el anillo de enteros algebraicos de $\mathbb{Q}(\omega)$. De momento observemos que, por la unicidad de la forma reducida, un elemento de $\mathbb{Z}[\omega]$ que sea racional es necesariamente un número entero (pues su forma reducida tiene que ser $\alpha = a_0$ por ser racional, y a_0 tiene que ser entero por estar en $\mathbb{Z}[\omega]$).

Teorema 17.9 *La norma es una aplicación* $\mathrm{N} : \mathbb{Q}(\omega) \longrightarrow \mathbb{Q}$ *que sólo toma valores positivos y cumple*

$$\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta).$$

Además, se restringe a una aplicación $\mathrm{N} : \mathbb{Z}[\omega] \longrightarrow \mathbb{N}$.

Demostración: Para probar que $\mathrm{N}(\alpha)$ es siempre un número racional, basta aplicar el teorema 17.7 (o el argumento particular que hemos dado previamente para el caso $p = 5$ tomando $g = 2$), teniendo en cuenta que

$$\sigma_g(\mathrm{N}(\alpha)) = \prod_{j\in U_p} \sigma_g(\sigma_j(\alpha)) = \prod_{j\in U_p} \sigma_{jg}(\alpha) = \prod_{j\in U_p} \sigma_j(\alpha) = \mathrm{N}(\alpha),$$

donde hemos usado que las clases $1\cdot g, \ldots, (p-1)\cdot g$ son todas las clases de U_p.

Es evidente que los conjugados de los enteros ciclotómicos son enteros ciclotómicos, luego la norma de un entero ciclotómico es a la vez un número racional

y un entero ciclotómico, y acabamos de observar que esto implica que es un número entero. Vamos a ver que la norma no toma valores negativos, y así en particular las normas de los enteros ciclotómicos serán números naturales. Lo probamos para $p = 5$ y dejamos al lector la adaptación mínima que requiere el argumento para el caso general:

Teniendo en cuenta que $\sigma_3 = \sigma_4\sigma_2$, y que σ_4 es la conjugación compleja, tenemos que

$$\mathrm{N}(\alpha) = \sigma_1(\alpha)\sigma_4(\alpha)\sigma_2(\alpha)\sigma_4(\sigma_2(\alpha)) = \alpha\bar{\alpha}\,\sigma_2(\alpha)\overline{\sigma_2(\alpha)} = |\alpha|^2|\sigma_2(\alpha)|^2 \geq 0.$$

Por último, que la norma conserva productos es consecuencia inmediata de que las conjugaciones lo hacen. ∎

La norma nos da una forma alternativa de calcular el inverso de un número ciclotómico (incluso de probar su existencia):

Para calcular el inverso de un número ciclotómico $\alpha \neq 0$ basta despejar $\alpha = \sigma_1(\alpha)$ en la definición de norma (teniendo en cuenta que, obviamente, la norma de un número no nulo es no nula):

$$\alpha^{-1} = \frac{\sigma_2(\alpha)\cdots\sigma_{p-1}(\alpha)}{\mathrm{N}(\alpha)}. \tag{17.3}$$

Ejemplos de cálculo de normas Calcular la norma de un número ciclotómico quinto requiere hacer tres multiplicaciones, lo cual es algo tedioso. Sin embargo, en algunas ocasiones podemos tomar atajos. Por ejemplo, consideremos la igualdad

$$(x-\omega)(x-\omega^2)\cdots(x-\omega^{p-1}) = x^{p-1} + x^{p-2} + \cdots + x + 1.$$

Haciendo $x = 1$ obtenemos:

$$(1-\omega)(1-\omega^2)\cdots(1-\omega^{p-2}) = p, \tag{17.4}$$

pero todos los factores del miembro izquierdo son conjugados, luego

$$\mathrm{N}(1-\omega^j) = p, \qquad j = 1, \ldots, p-1. \tag{17.5}$$

Ahora consideramos la igualdad

$$(t-1)(t-\omega)(t-\omega^2)\cdots(t-\omega^{p-1}) = t^p - 1.$$

Haciendo $t = x/y$ y multiplicando por y^p queda:

$$(x+y)(x+y\omega)(x+y\omega^2)\cdots(x+y\omega^{p-1}) = x^p + y^p.$$

Si $x \neq -y$ son números racionales, todos los factores del miembro izquierdo menos el primero son conjugados, por lo que tenemos la fórmula siguiente para la norma de un binomio:

$$\mathrm{N}(x+y\omega^j) = \frac{x^p + y^p}{x+y}, \tag{17.6}$$

Si $x = -y$ tenemos que

$$\mathrm{N}(x - x\omega^k) = \mathrm{N}(x)\,\mathrm{N}(1 - \omega^k) = px^{p-1}.$$

Ejercicio: Calcular la norma de los enteros de orden 5 siguientes:

$$\omega, \quad \omega + 2, \quad \eta = \omega^4 + \omega = \omega(\omega^3 + 1), \quad \eta' = \omega^3 + \omega^2 = \omega^2(\omega + 1).$$

En el caso $p = 5$, para calcular la norma de un número ciclotómico arbitrario podemos simplificar el proceso expresando la norma en la forma

$$\mathrm{N}(\alpha) = \alpha\sigma_4(\alpha)\sigma_2(\alpha)\sigma_2(\sigma_4(\alpha)) = \alpha\bar{\alpha}\sigma_2(\alpha\bar{\alpha}).$$

Así, en primer lugar calculamos $\alpha\bar{\alpha}$, que es un número ciclotómico real, luego el resultado será de la forma $\alpha\bar{\alpha} = a + b\eta$ (o $a + b\eta'$, según resulte más cómodo para los cálculos), y en segundo lugar observamos que $\sigma_2(a + b\eta') = a + b\eta$, pues $\sigma_2(\eta') = \sigma_2(\omega^3 + \omega^2) = \omega^4 + \omega = \eta$. Por lo tanto

$$\mathrm{N}(\alpha) = (a + b\eta')(a + b\eta) = a^2 + b^2\eta\eta' + ab(\eta + \eta') = a^2 - ab + b^2,$$

donde hemos usado que $\eta + \eta' = \eta\eta' = -1$.

Ejercicio: Usar el procedimiento que acabamos de describir para calcular el valor de $\mathrm{N}(-\omega^3 + \omega^2 - \omega)$.

Un poco más en general, si queremos calcular, por ejemplo,

$$\frac{\omega^2 + 2}{\omega + 2},$$

podemos hacerlo multiplicando el numerador y el denominador por los tres conjugados no triviales del denominador:

$$\frac{\omega^2 + 2}{\omega + 2} = \frac{(\omega^2 + 2)(\omega^4 + 2)(\omega^3 + 2)(\omega^2 + 2)}{\mathrm{N}(\omega + 2)}.$$

La fórmula (17.6) nos da que

$$\mathrm{N}(\omega + 2) = \frac{2^5 + 1^5}{2 + 1} = 11.$$

Por lo tanto:

$$\frac{\omega^2 + 2}{\omega + 2} = \frac{(-2\omega^3 - \omega + 2)(2\omega^3 + 2\omega^2 + 5)}{11} = \frac{-6\omega^3 + 6\omega^2 - 7\omega + 8}{11}.$$

■

La medida de un entero ciclotómico Terminamos esta sección asociando otro número racional a cada número ciclotómico quinto, que nos proporcionará una cota para su norma. Gauss lo llamó la *medida* de un número ciclotómico α, definida como

$$M(\alpha) = |\sigma_1(\alpha)|^2 + |\sigma_2(\alpha)|^2 + |\sigma_3(\alpha)|^2 + |\sigma_4(\alpha)|^2.$$

No es difícil probar que es un número racional mostrando que es invariante por σ_2, pero vamos a necesitar una expresión explícita que volverá este hecho inmediato. Ante todo, como σ_1 y σ_4 son conjugados complejos, al igual que σ_2 y σ_3, ambos pares de números tienen el mismo valor absoluto, luego

$$M(\alpha) = 2|\alpha|^2 + 2|\sigma_2(\alpha)|^2 = 2(\alpha\bar{\alpha} + \sigma_2(\alpha)\sigma_2(\bar{\alpha})) = 2(\alpha\bar{\alpha} + \sigma_2(\alpha\bar{\alpha})).$$

Si $\alpha = a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0$, entonces $\bar{\alpha} = \sigma_4(\alpha) = a_1\omega^4 + a_2\omega^3 + a_3\omega^2 + a_0$, luego $\alpha\bar{\alpha}$ se calcula así:

$$\begin{array}{ccccc}
0 & a_3 & a_2 & a_1 & a_0 \\
a_1 & a_2 & a_3 & 0 & a_0 \\
\hline
0 & a_0a_3 & a_0a_2 & a_0a_1 & a_0^2 \\
a_3a_2 & a_3a_1 & a_3a_0 & 0 & a_3^2 \\
a_2a_1 & a_2a_0 & 0 & a_2a_3 & a_2^2 \\
a_1a_0 & 0 & a_1a_3 & a_1a_2 & a_1^2 \\
\hline
B & C & C & B & A
\end{array}$$

donde

$$A = a_0^2 + a_1^2 + a_2^2 + a_3^2, \quad B = a_0a_1 + a_1a_2 + a_2a_3, \quad C = a_0a_2 + a_0a_3 + a_1a_3.$$

Equivalentemente: $\alpha\bar{\alpha} = A + B\eta + C\eta'$, de donde

$$\begin{aligned} M(\alpha) &= 2(A + B\eta + C\eta') + 2(A + B\eta' + C\eta) = 4A - 2(B + C) \\ &= 5(a_0^2 + a_1^2 + a_2^2 + a_3^2) - (a_0 + a_1 + a_2 + a_3)^2. \end{aligned}$$

Esta expresión muestra que la medida de un número ciclotómico es un número racional y que la medida de un entero ciclotómico es un número natural (notemos que la medida es no negativa por definición). En particular nos va a interesar la cota

$$M(\alpha) \le 5(a_0^2 + a_1^2 + a_2^2 + a_3^2). \tag{17.7}$$

La relación entre la norma y la medida viene dada por el teorema siguiente:

Teorema 17.10 *Si α es un número ciclotómico de orden 5, entonces*

$$\mathrm{N}(\alpha) \le \left(\frac{M(\alpha)}{4}\right)^2.$$

Demostración: Tenemos que

$$\mathrm{N}(\alpha)^2 = |\,\mathrm{N}(\alpha)|^2 = |\sigma_1(\alpha)|^2|\sigma_2(\alpha)|^2|\sigma_3(\alpha)|^2|\sigma_4(\alpha)|^2$$

Por la desigualdad entre la media aritmética y la geométrica [ITAn 4.9]:

$$\sqrt{\mathrm{N}(\alpha)} \le \frac{|\sigma_1(\alpha)|^2 + |\sigma_2(\alpha)|^2 + |\sigma_3(\alpha)|^2 + |\sigma_4(\alpha)|^2}{4} = \frac{M(\alpha)}{4}.$$

■

17.5 La traza

Vamos a demostrar que el anillo de enteros algebraicos de $\mathbb{Q}(\omega)$ es precisamente $\mathbb{Z}[\omega]$, y por ello en adelante nos referiremos a sus elementos como "enteros ciclotómicos". Para eso usaremos un concepto similar al de la norma:

Definición 17.11 Definimos la *traza* como la aplicación $\mathrm{Tr} : \mathbb{Q}(\omega) \longrightarrow \mathbb{Q}$ dada por

$$\mathrm{Tr}(\alpha) = \sum_{j=1}^{p-1} \sigma_j(\alpha).$$

Exactamente el mismo argumento empleado con la norma prueba que la traza de un número ciclotómico es invariante por todas las conjugaciones, por lo que es un número racional. Otras propiedades obvias de la traza son:

$$\mathrm{Tr}(\alpha + \beta) = \mathrm{Tr}(\alpha) + \mathrm{Tr}(\beta),$$

así como que si $a \in \mathbb{Q}$, entonces $\mathrm{Tr}(a\alpha) = a\,\mathrm{Tr}(\alpha)$.

Pero la propiedad esencial que vamos a aprovechar es que si $\alpha \in \mathbb{Q}(\omega)$ es un entero algebraico, entonces $\mathrm{Tr}(\alpha), \mathrm{N}(\alpha) \in \mathbb{Z}$.

En efecto, si α es un entero algebraico, esto significa que $p(\alpha) = 0$, donde $p(x) \in \mathbb{Z}[x]$ es un polinomio mónico, pero entonces, como las conjugaciones σ_j conservan sumas y productos y dejan invariantes a los números racionales, se cumple también que $p(\sigma_j(\alpha)) = 0$, luego $\sigma_j(\alpha)$ también es un entero algebraico.

Por ejemplo, si se cumple que $\alpha^4 - 3\alpha^3 + 2\alpha - 7 = 0$, entonces también $\sigma_j(\alpha)^4 - 3\sigma_j(\alpha)^3 + 2\sigma_j(\alpha) - 7 = 0$.

Por consiguiente, si $\alpha \in \mathbb{Q}(\omega)$ es un entero algebraico, entonces $\mathrm{Tr}(\alpha), \mathrm{N}(\alpha)$ son enteros algebraicos (por ser suma o producto de enteros algebraicos) y números racionales, luego son enteros racionales.

Así ya podemos probar:

Teorema 17.12 *Si p es un primo impar, el anillo de enteros algebraicos del cuerpo ciclotómico $\mathbb{Q}(\omega)$ de orden p es $\mathbb{Z}[\omega]$.*

Demostración: Llamemos $\mathcal{O}$ al anillo de enteros algebraicos de $\mathbb{Q}(\omega)$. Como ω es ciertamente un entero algebraico y al operar enteros algebraicos obtenemos enteros algebraicos, es inmediato que $\mathbb{Z}[\omega] \subset \mathcal{O}$. Tenemos que probar la implicación opuesta.

Sea $\alpha = \sum_{i=0}^{p-2} a_i\omega^i \in \mathcal{O}$. Hemos de probar que todos los coeficientes son enteros racionales. En principio sabemos que la traza es un entero. Más aún, para cada $0 \le k \le p-2$ tenemos que $\mathrm{Tr}(\alpha\omega^{-k}) \in \mathbb{Z}$. Así tenemos la misma información sobre todos los coeficientes:

$$\mathrm{Tr}(\alpha\omega) = -\sum_{i=0}^{p-2} a_i \in \mathbb{Z}, \qquad \mathrm{Tr}(\alpha\omega^{-k}) = pa_k - \sum_{i=0}^{p-2} a_i \in \mathbb{Z}, \qquad \text{si } k \neq p-1$$

Por lo tanto $pa_k \in \mathbb{Z}$ para todo $k = 0, \dots, p-1$. Llamemos $b_k = pa_k$. Hemos de probar que $p \mid b_k$ para todo k, con lo que los a_k serán también enteros. Consideremos $\lambda = 1 - \omega$. Sustituyendo $\omega = 1 - \lambda$ y desarrollando:

$$p\alpha = \sum_{i=0}^{p-2} b_i \omega^i = \sum_{i=0}^{p-2} c_i \lambda^i, \qquad \text{donde } c_i = \sum_{j=i}^{p-2} (-1)^i \binom{j}{i} b_j \in \mathbb{Z},$$

Como $\lambda = 1 - \omega$, por simetría se cumple también

$$b_i = \sum_{j=i}^{p-2} (-1)^i \binom{j}{i} c_j, \qquad i = 0, \dots, p-2.$$

Por lo tanto basta probar que $p \mid c_j$ para todo j, pues entonces estas fórmulas implican que p también divide a los b_i. Lo probaremos por inducción. Suponemos que $p \mid c_i$ para cada $i \leq k-1$ y vamos a probar que $p \mid c_k$, donde $0 \leq k \leq p-2$. En la sección anterior hemos visto que

$$\mathrm{N}(\lambda) = \mathrm{N}(1-\omega) = (1-\omega)\cdots(1-\omega^{p-1}) = p.$$

Más aún, cada factor se descompone como

$$1 - \omega^i = (1-\omega)(\omega^{i-1} + \cdots + \omega + 1),$$

por lo que $p = \delta\lambda^{p-1}$, para cierto $\delta \in \mathcal{O}$. Esto implica que $p \equiv 0 \pmod{\lambda^{p-1}}$. Por otro lado,

$$0 \equiv p\alpha = \sum_{i=0}^{p-2} c_i \lambda^i \equiv c_k \lambda^k \pmod{\lambda^{k+1}},$$

pues los términos anteriores a $c_k\lambda^k$ son múltiplos de p por hipótesis de inducción y los posteriores son múltiplos de λ^{k+1} directamente.

Esto equivale a la relación $c_k\lambda^k = \eta\lambda^{k+1}$ para un cierto $\eta \in \mathcal{O}$, luego $c_k = \eta\lambda$. Finalmente tomamos normas: $c_k^{p-1} = \mathrm{N}(c_k) = \mathrm{N}(\eta)\,\mathrm{N}(\lambda) = p\,\mathrm{N}(\eta)$, luego en efecto $p \mid c_k$. ∎

17.6 La aritmética de los enteros ciclotómicos

Tras haber estudiado el álgebra básica de los enteros ciclotómicos, ya estamos en condiciones de estudiar su aritmética, al menos en el caso $p = 5$. El resultado fundamental es que los enteros ciclotómicos (de orden 5) son un dominio euclídeo y, por consiguiente, son un dominio de factorización única:

Teorema 17.13 *El anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos de orden 5 es un dominio euclídeo.*

Demostración: Más concretamente, vamos a probar que $\mathbb{Z}[\omega]$ es un dominio euclídeo tomando como norma euclídea la norma algebraica.

El mismo argumento empleado al principio de la sección 9.4 muestra que basta probar que, para todo número ciclotómico α, existe un entero ciclotómico γ tal que $\mathrm{N}(\alpha-\gamma)<1$.

En efecto,[3] si se cumple esto, para dividir un dividendo Δ entre un divisor no nulo δ, basta encontrar un γ tal que $\mathrm{N}(\Delta/\delta-\gamma)<1$, pues entonces, llamando $\rho=\Delta-\delta\gamma$, se cumple que $\Delta=\delta\gamma+\rho$, con

$$\mathrm{N}(\rho)=\mathrm{N}(\delta)\,\mathrm{N}(\Delta/\delta-\gamma)<\mathrm{N}(\delta).$$

Sea, pues, $\alpha=a_3\omega^3+a_2\omega^2+a_1\omega+a_0$ y empezamos descomponiendo[4]

$$\begin{aligned}
4a_0-a_1-a_2-a_3 &= e_0+r_0\\
4a_1-a_0-a_2-a_3 &= e_1+r_1\\
4a_2-a_0-a_1-a_3 &= e_2+r_2\\
4a_3-a_0-a_1-a_2 &= e_3+r_3\\
-a_0-a_1-a_2-a_3 &= e_4+r_4
\end{aligned}$$

donde los e_i son enteros y $0\le r_i<1$. Así:

$$\begin{aligned}
\alpha &= \frac{e_3-e_4+r_3-r_4}{5}\,\omega^3+\frac{e_2-e_4+r_2-r_4}{5}\,\omega^2\\
&+ \frac{e_1-e_4+r_1-r_4}{5}\,\omega+\frac{e_0-e_4+r_0-r_4}{5}
\end{aligned}$$

y además $e_0+e_1+e_2+e_3+e_4+r_0+r_1+r_2+r_3+r_4=0$.

Consideremos en primer lugar el caso en que $r_0+r_1+r_2+r_3+r_4=0$. Como todos los sumandos son mayores o iguales que 0, esto equivale a que $r_i=0$ para todo i. En particular

$$a_i=\frac{e_i-e_4}{5}.$$

Tomamos números enteros tales que $|a_i-n_i|\le 1/2$, es decir, $|e_i-e_4-5n_i|\le 5/2$, pero podemos rebajar la cota hasta 2, porque el miembro izquierdo es entero, luego de hecho tenemos que $|a_i-n_i|\le 2/5$.

[3]Por ejemplo, si queremos calcular un cociente y un resto de la división de $\Delta=\omega^2+2$ entre $\delta=\omega+2$, recordamos que ya habíamos calculado

$$\frac{\Delta}{\delta}=\alpha=-\frac{6}{11}\omega^3+\frac{6}{11}\omega^2-\frac{7}{11}\omega+\frac{8}{11}.$$

Ahora tenemos que aproximar este cociente por un entero ciclotómico. Vamos a tratar este ejemplo en particular paralelamente al caso general.

[4]En el ejemplo es fácil ver que

$$e_0=3, r_0=\frac{6}{11},\quad e_1=-4, r_1=\frac{8}{11},\quad e_2=2, r_2=\frac{7}{11},$$

$$e_3=-3, r_3=\frac{2}{11},\quad e_4=-1, r_4=\frac{10}{11}.$$

Definimos $\gamma = n_3\omega^3 + n_2\omega^2 + n_1\omega + n_0$, que cumple que

$$N(\alpha-\gamma) \le \left(\frac{M(\alpha-\gamma)}{4}\right)^2$$

$$\le \left(\frac{5(|a_0-n_0|^2+|a_1-n_1|^2+|a_2-n_2|^2+|a_3-n_3|^2)}{4}\right)^2 \le \left(\frac{4}{5}\right)^2 < 1,$$

donde hemos usado el teorema 17.10 que relaciona la norma y la medida, así como la desigualdad (17.7). Esto termina la prueba en el caso en que todos los r_i son nulos. Llamemos ahora

$$m = -e_0 - e_1 - e_2 - e_3 - e_4 = r_0 + r_1 + r_2 + r_3 + r_4 > 0.$$

Notemos que la primera expresión muestra que m es un número entero,[5] mientras que la segunda implica que $0 < m < 5$, luego $m = 1, 2, 3, 4$. Además $e_0 + e_1 + e_2 + e_3 + e_4 + m = 0$, luego

$$e_0 - e_4 + e_1 - e_4 + e_2 - e_4 + e_3 - e_4 + m = -5e_4 \equiv 0 \text{ (mód 5)}.$$

Todo número entero es congruente módulo 5 con uno de los números $-2, -1, 0, 1, 2$, luego podemos tomar enteros s_i tales que[6] $e_i - e_4 \equiv s_i$ (mód 5) y $|s_i| \le 2$. La congruencia anterior equivale entonces a que

$$s_0 + s_1 + s_2 + s_3 \equiv -m \text{ (mód 5)}. \tag{17.8}$$

Supongamos ahora que algún $s_i = 0$. Digamos, por ejemplo, que $s_2 = 0$, con lo que $e_2 - e_4 = 5k_2$. Tomemos como antes enteros n_i tales que $|a_i - n_i| \le 1/2$. Entonces, de hecho,

$$|a_2 - n_2| = \left|\frac{e_2 - e_4 + r_2 - r_4}{5} - n_2\right| = \left|\frac{r_2 - r_4}{5} + k_2 - n_2\right| = \left|\frac{r_2 - r_4}{5}\right| < \frac{1}{5}.$$

La última igualdad se debe a que el miembro izquierdo debe ser $\le 1/2$, para lo cual es necesario que $k_2 - n_2 = 0$, o de lo contrario dicho miembro izquierdo sería al menos $4/5$. Por lo tanto, si llamamos

$$\gamma' = n_3\omega^3 + n_2\omega^2 + n_1\omega + n_0,$$

se trata de un entero ciclotómico tal que

$$\alpha - \gamma' = c_3\omega^3 + c_2\omega^2 + c_1\omega + c_0,$$

donde $|c_i| = |a_i - n_i| \le 1/2$, pero para uno de los índices se cumple, más precisamente, $|c_i| < 1/5$. Con esto ya podríamos llegar a la conclusión, pero en vez de esto vamos a considerar ahora el caso que tenemos pendiente y veremos que el razonamiento final vale para ambos casos.

[5] En nuestro ejemplo es $m = 3$.

[6] En nuestro ejemplo $s_0 = -1$, $s_1 = 2$, $s_2 = -2$, $s_3 = -2$.

Ahora suponemos que ningún $s_i = 0$. Esto significa que los cuatro s_i toman los valores $-2, -1, 1, 2$, y tiene que haber al menos dos iguales, digamos $s_i = s_j$, pues en caso contrario los cuatro s_i tomarían los cuatro valores posibles y su suma sería 0, en contradicción con (17.8), ya que $m = 1, 2, 3, 4$.

Supongamos, por ejemplo, que es $s_2 = s_3$, con lo que $e_2 \equiv e_3 \pmod 5$. Entonces[7]

$$\begin{aligned}\omega\alpha &= \frac{e_3 - e_4 + r_3 - r_4}{5}\,\omega^4 + \frac{e_2 - e_4 + r_2 - r_4}{5}\,\omega^3\\ &+ \frac{e_1 - e_4 + r_1 - r_4}{5}\,\omega^2 + \frac{e_0 - e_4 + r_0 - r_4}{5}\omega\\ &= \frac{e_2 - e_3 + r_2 - r_3}{5}\,\omega^3 + \frac{e_1 - e_3 + r_1 - r_3}{5}\,\omega^2\\ &+ \frac{e_0 - e_3 + r_0 - r_3}{5}\,\omega + \frac{e_4 - e_3 + r_4 - r_3}{5}\end{aligned}$$

y así $\omega\alpha$ está en el caso anterior, porque $e_2 - e_3$ es múltiplo de 5. (En general, si $s_i = s_j$, multiplicamos α por la potencia ω^k adecuada para que $e_i - e_4$ aparezca acompañando a ω^4, y así al reducir la expresión aparece $e_j - e_i$ acompañando a alguna potencia de ω.)

Así, en cualquiera de los dos casos que tenemos pendientes, existe un índice k y un entero ciclotómico γ' de manera que[8]

$$\omega^k\alpha - \gamma' = c_3\omega^3 + c_2\omega^2 + c_1\omega + c_0,$$

con $|c_i| \le 1/2$ para $i = 0, 1, 2, 3$, pero de modo que uno de los coeficientes cumple, de hecho, $|c_i| < 1/5$. Por consiguiente,

$$M(\omega^k\alpha - \gamma') \le 5\left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{25}\right) = \frac{79}{20},$$

$$N(\alpha - \omega^{-k}\gamma') = N(\omega^{-k})N(\omega^k\alpha - \gamma') = N(\omega^k\alpha - \gamma) \le \left(\frac{79}{80}\right)^2 < 1,$$

luego $\gamma = \omega^{-k}\gamma'$ cumple lo requerido.[9] ∎

Ahora podemos dar una prueba directa del teorema 17.12 para el caso $p = 5$:

Teorema 17.14 *El anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos de orden 5 es el anillo de enteros algebraicos del cuerpo $\mathbb{Q}(\omega)$.*

---

[7] En nuestro ejemplo,

$$\omega\alpha = \frac{12}{11}\omega^3 - \frac{1}{11}\omega^2 + \frac{14}{11}\omega + \frac{6}{11}.$$

[8] En nuestro ejemplo, redondeando los coeficientes que aparecen en la not al pie precedente, obtenemos los valores $c_0 = 1$, $c_1 = 1$, $c_2 = 0$, $c_3 = 1$, luego $\gamma' = \omega^3 + \omega + 1$.

[9] En nuestro ejemplo, $\gamma = \omega^{-1}\gamma' = \omega^4 + \omega^2 + 1 = -\omega^3 - \omega$. Éste es el cociente de la división euclídea. El resto es $\rho = \omega^2 + 2 - (\omega + 1)(-\omega^3 - \omega) = \omega^3 + \omega^2 + \omega + 1 = -\omega^4$, donde $\mathrm{N}(\rho) = 1 < 11 = \mathrm{N}(\delta)$.

DEMOSTRACIÓN: Como ω es claramente un entero algebraico, todos los elementos de $\mathbb{Z}[\omega]$ lo son (por el teorema 8.15). Es claro que $\mathbb{Q}(\omega)$ es el cuerpo de cocientes de $\mathbb{Z}[\omega]$ y, si $\alpha \in \mathbb{Q}(\omega)$ es un entero algebraico, entonces es raíz de un polinomio con coeficientes en $\mathbb{Z}$, luego en particular en $\mathbb{Z}[\omega]$, y el teorema 2.28 implica que $\alpha \in \mathbb{Z}[\omega]$. ∎

Para analizar las descomposiciones en factores primos de los enteros ciclotómicos necesitamos reconocer las unidades. El anillo de los enteros ciclotómicos de orden 3 no son sino el anillo de los enteros de Eisenstein, y sabemos que sus unidades son $\pm 1, \pm\zeta, \pm\zeta^2$. Similarmente, el anillo de los enteros ciclotómicos de orden 4 no es sino el anillo de los enteros de Gauss, y sabemos que sus unidades son $\pm 1\ \pm i$, por lo que cabría esperar que las unidades ciclotómicas (de orden 5) fueran $\pm 1, \pm\omega, \pm\omega^2, \pm\omega^3, \pm\omega^4$. Sin embargo, no es así, pues hemos visto que el anillo $\mathbb{Z}[\omega]$ contiene al anillo de enteros $\mathbb{Z}[\epsilon]$ del cuerpo cuadrático $\mathbb{Q}\big(\sqrt{5}\,\big)$, y $\mathbb{Z}[\epsilon]$ tiene infinitas unidades, que obviamente también son unidades ciclotómicas. Concretamente, sabemos[10] que las unidades de $\mathbb{Z}[\epsilon]$ son los enteros cuadráticos de la forma $\pm\epsilon^n$.

Necesitaremos determinar cuáles son exactamente las unidades ciclotómicas de orden 5, pero de momento observemos que podemos reconocerlas a través de su norma, como en el caso de los cuerpos cuadráticos:

Teorema 17.15 *Un entero ciclotómico α es una unidad de $\mathbb{Z}[\omega]$ si y sólo si $\mathrm{N}(\alpha) = 1$.*

DEMOSTRACIÓN: Si α es una unidad, entonces α^{-1} es también un entero ciclotómico, y se cumple que $\mathrm{N}(\alpha)\,\mathrm{N}(\alpha^{-1}) = \mathrm{N}(1) = 1$ y, como las normas son números naturales, $\mathrm{N}(\alpha) = 1$. Recíprocamente, si $\mathrm{N}(\alpha) = 1$, la fórmula (17.3) muestra que $\alpha^{-1} \in \mathbb{Z}[\omega]$, luego α es una unidad. ∎

Como en el caso cuadrático, todo primo ciclotómico divide a un primo racional. Más aún:

Teorema 17.16 *Todo primo ciclotómico π divide a un único primo racional p. Además, todos los factores primos de p tienen la misma norma, luego $\mathrm{N}(\pi) = p^i$, para un cierto exponente $i \mid p-1$.*

DEMOSTRACIÓN: Por definición de norma, $\pi \mid \mathrm{N}(\pi)$, pero $\mathrm{N}(\pi)$ es un entero racional (que no puede ser 1, pues entonces π sería una unidad), luego podemos descomponerlo en producto de primos racionales y, por ser primo, π tiene que dividir a alguno de ellos, digamos p.

En general, si π divide a un primo racional p, entonces $\mathrm{N}(\pi) \mid \mathrm{N}(p) = p^{p-1}$, luego $\mathrm{N}(\pi) = p^i$, con $1 \le i \le p-1$. Esto prueba que π divide a un único primo racional p (el único que divide a $\mathrm{N}(\pi)$).

Si p tiene r factores primos y probamos que todos los factores primos de p tienen la misma norma, digamos p^i, entonces $p^{p-1} = \mathrm{N}(p) = p^{ri}$, luego $i \mid p-1$.

[10] Véase el ejemplo tras la definición 9.7, en el cual llamamos ω al número que aquí llamamos ϵ.

En efecto, si π y π' son divisores primos de p, entonces $\mathrm{N}(\pi') = p^i$, luego $\pi \mid \mathrm{N}(\pi') = \sigma_1(\pi)\cdots\sigma_{p-1}(\pi)$, luego existe un j tal que $\pi \mid \sigma_j(\pi')$. Pero es fácil ver que $\sigma_j(\pi')$ cumple la definición de primo, luego π y $\sigma_j(\pi')$ son asociados, es decir, $\pi = \delta\sigma_j(\pi')$, para cierta unidad δ, luego

$$\mathrm{N}(\pi) = \mathrm{N}(\delta)\,\mathrm{N}(\sigma_j(\pi')) = \mathrm{N}(\sigma_j(\pi')) = \mathrm{N}(\pi'),$$

pues ambas normas son el producto de los conjugados de π'. ∎

La factorización de p El único primo ciclotómico con el que necesitaremos trabajar es $\lambda = \omega - 1$, que resulta ser primo incluso si $\mathbb{Z}[\omega]$ no tiene factorización única.[11] En efecto:

Teorema 17.17 *En el anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos de orden primo p, se cumple que $\lambda = \omega - 1$ es primo, que $p = \delta\lambda^{p-1}$, para cierta unidad δ y que el anillo de clases de restos $\mathbb{Z}[\omega]/(\lambda)$ es $\mathbb{Z}_p$.*

Demostración: En la prueba del teorema 17.12 hemos visto que $\mathrm{N}(\lambda) = p$, así como que $p = \delta\lambda^{p-1}$, para cierto entero δ que tiene que tener norma 1, luego es una unidad. Basta probar que $\mathbb{Z}[\omega]/(\lambda)$ es $\mathbb{Z}_p$, pues entonces tendremos que se trata de un dominio íntegro, luego el teorema 13.13 (y el ejercicio previo) nos dan que λ es primo.

En efecto, se cumple que $\omega \equiv 1$ (mód λ), de donde se sigue que todo entero ciclotómico es congruente con un entero módulo λ. Concretamente:

$$a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0 \equiv a_{p-1} + \cdots + a_1 + a_0 \text{ (mód } \lambda),$$

y, a su vez, dividiendo euclídeamente la suma final entre p, concuimos que

$$\mathbb{Z}[\omega]/(\lambda) = \{\bar{0}, \ldots, \overline{p-1}\}.$$

Además, estas p clases de restos son distintas entre sí, pues si $\lambda \mid a - b$, con $0 \le a, b < p$, entonces $p = \mathrm{N}(\lambda) \mid \mathrm{N}(a-b) = (a-b)^{p-1}$, luego $p \mid a - b$, luego $a = b$.

Es claro entonces que el anillo cociente $\mathbb{Z}[\omega]/(\lambda)$ es exactamente lo mismo que $\mathbb{Z}_p$ (con las mismas operaciones), luego es un cuerpo. ∎

En el caso $p = 5$, a partir de (17.4) obtenemos, concretamente, que

$$5 = \lambda\,(\omega+1)\lambda\,(\omega^2+\omega+1)\lambda\,(\omega^3+\omega^2+\omega+1)\lambda,$$

y operando:

$$5 = (\omega^3 - \omega - 1)\lambda^4.$$

∎

[11] En el caso $p = 5$ basta observar que, por (17.5), sabemos que $\mathrm{N}(\lambda) = p$, de donde se sigue fácilmente que λ es irreducible, y por la factorización única tiene que ser primo.

Unidades ciclotómicas Vamos a determinar exactamente cuáles son las unidades ciclotómicas de orden 5. Sabemos que hay infinitas, porque incluyen a todas las unidades $\pm\epsilon^n$ del anillo $\mathbb{Z}[\epsilon]$ de los enteros algebraicos de $\mathbb{Q}(\sqrt{5})$. Vamos a probar que no hay muchas más:

Teorema 17.18 *Las unidades del anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos de orden 5 son los enteros ciclotómicos de la forma $\pm\omega^m\epsilon^n$, donde $\epsilon = (1+\sqrt{5})/2$, n es un entero racional y $m = 0, 1, 2, 3, 4$.*

Sabemos que los números de la forma $\pm\epsilon^n$ son las unidades del anillo $\mathbb{Z}[\epsilon]$ de los enteros ciclotómicos reales, luego el teorema anterior equivale a que toda unidad ciclotómica es el producto de una potencia de ω por una unidad ciclotómica real. Así pues, basta demostrar el teorema siguiente, que es válido para p arbitrario:

Teorema 17.19 (Lema de Kummer) *Si α es una unidad ciclotómica, existe una unidad ciclotómica real β tal que $\alpha = \omega^m\beta$, para cierto $m = 0, \ldots, p-1$.*

Ejemplo Consideremos la unidad $\alpha = \omega^3 - \omega - 1$ que nos ha aparecido en la factorización de 5 en el anillo de enteros ciclotómicos de orden 5. Probando obtenemos que hay que multiplicarla por ω^2 para obtener un entero real, luego multiplicando por $\omega^3\omega^2 = 1$ obtenemos:

$$\alpha = \omega^3(-\omega^3 - \omega^2 + 1) = \omega^3(-\eta' + 1) = \omega^3\epsilon^2.$$

Similarmente,

$$\omega + 1 = \omega^3(\omega^3 + \omega^2) = \omega^3\eta' = -\omega^3\epsilon.$$
$$\omega^2 + \omega + 1 = \omega(\omega^4 + \omega + 1) = \omega(-\omega^3 - \omega^2) = \omega\epsilon. \qquad \blacksquare$$

DEMOSTRACIÓN: Es claro que los conjugados de una unidad son unidades. En particular, el conjugado complejo $\bar{\alpha} = \sigma_{-1}(\alpha)$ de la unidad dada es también una unidad, luego también lo es el cociente $\alpha/\bar{\alpha}$, que podemos expresar en forma canónica:

$$\frac{\alpha}{\bar{\alpha}} = a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0.$$

Consideramos el polinomio

$$E(x) = a_{p-1}x^{p-1} + \cdots + a_1x + a_0.$$

Entonces

$$\frac{\bar{\alpha}}{\alpha} = \overline{E(\omega)} = E(\omega^{p-1}).$$

Se cumple, pues, que

$$E(\omega)E(\omega^{p-1}) = \frac{\alpha}{\bar{\alpha}}\frac{\bar{\alpha}}{\alpha} = 1.$$

Dividimos

$$E(x)E(x^x) = (x^p - 1)Q(x) + R(x), \tag{17.9}$$

donde el resto tiene grado menor que p, es decir,

$$R(x) = b_{p-1}x^{p-1} + \cdots + b_1x + b_0.$$

Los monomios de $E(x)$ son de la forma $a_i x^i$, mientras que los de $E(x^{p-1})$ son de la forma $a_j x^{(p-1)j}$. Para multiplicar $E(x)$ por $E(x^{p-1})$ tenemos que multiplicar cada monomio de $E(x)$ por cada monomio de $E(x^{p-1})$ y luego sumar todos los productos. Dichos productos son de la forma

$$a_i x^{(p-1)i} a_j x^j = a_i a_j x^{pi+j-i} = a_i a_j x^{pc_{ij}+r_{ij}},$$

donde hemos expresado $pj + i - j = pc_{ij} + r_{ij}$, con $0 \le r_{ij} < p$. Tomando congruencias en la igualdad precedente vemos que $r_{ij} \equiv i - j \pmod{p}$. Observemos también que $c_{ij} \ge 0$, pues $pc_{ij} + r_{ij} = (p-1)j + i \ge 0$ y si fuera $c_{ij} < 0$ el primer miembro sería negativo. Seguimos operando:

$$a_i x^{(p-1)i} a_j x^j = a_i a_j x^{r_{ij}} \cdot x^{pc_{ij}} = a_i a_j x^{r_{ij}} (x^{pc_{ij}} - 1) + a_i a_j x^{r_{ij}}.$$

Ahora distinguimos tres casos:

1. Si $c_{ij} > 1$ usamos que $x^{c_{ij}-1} = (x-1)(x^{c_{ij}-1} + \cdots + x + 1)$. Sustituimos x por x^p y queda

   $$x^{pc_{ij}} - 1 = (x^p - 1)(x^{p(c_{ij}-1)} + \cdots + x^p + 1).$$

   Llamamos $Q_{ij}(x) = a_i a_j x^{r_{ij}} (x^{p(c_{ij}-1)} + \cdots + x^p + 1)$ y concluimos que

   $$a_i x^{(p-1)i} a_j x^j = Q_{ij}(x)(x^p - 1) + a_i a_j x^{r_{ij}},$$

   donde $Q_{ij}(x)$ es un polinomio con coeficientes enteros.

2. Si $c_{ij} = 1$ llegamos a esta misma expresión tomando $Q_{ij}(x) = a_i a_j x^{r_{ij}}$.

3. Si $c_{ij} = 0$ llegamos a esta misma expresión tomando $Q_{ij}(x) = 0$.

En los tres casos $Q_{ij}(x)$ es un polinomio con coeficientes enteros. Si ahora sumamos todos los productos de monomios, obtenemos que

$$E(x^{p-1})E(x) = \Big(\sum_{i,j=0}^{p-1} Q_{ij}(x)\Big)(x^p - 1) + \sum_{i,j=0}^{p-1} a_i a_j x^{r_{ij}}.$$

El último término es un polinomio de grado menor que p luego, por la unicidad de la división euclídea,

$$Q(x) = \sum_{i,j=0}^{p-1} Q_{ij}(x), \qquad R(x) = \sum_{i,j=0}^{p-1} a_i a_j x^{r_{ij}}.$$

Agrupando en $R(x)$ las potencias con el mismo exponente queda:

$$R(x) = \sum_{r=0}^{p-1} \Big(\sum_{r_{ij}=r} a_i a_j\Big) x^r.$$

Más precisamente, en el sumatorio interno, i, j recorren los índices entre 0 y $p-1$ que hacen que $r_{ij} = r$, pero recordemos que r_{ij} es simplemente el resto de $i - j$ módulo p.

Ahora bien, $i-j \equiv 0 \pmod{p}$ sólo se cumple cuando $i=j$, luego el término independiente del resto es

$$b_0 = \sum_{i=j} a_i a_j = a_0^2 + \cdots + a_{p-1}^2. \tag{17.10}$$

Por otra parte, haciendo $x=1$ en (17.9) resulta que $E(1)^2 = R(1)$, es decir,

$$(a_0 + a_1 + \cdots + a_{p-1})^2 = b_0 + b_1 + \cdots + b_{p-1}.$$

Haciendo $x=\omega$ en (17.9) obtenemos $1 = E(\omega)E(\omega^{p-1}) = R(\omega)$, luego

$$R(\omega) = b_{p-1}\omega^{p-1} + \cdots + b_1\omega + b_0 = 1.$$

Si llamamos $k = b_{p-1}$ y ponemos el miembro izquierdo en forma reducida, queda

$$(b_{p-2}-k)\omega^{p-2} + \cdots + (b_1-k)\omega + b_0 - k = 1.$$

Por la unicidad de la forma reducida concluimos que $b_i = k$ para $i = 1, \ldots, p-2$ y que $b_0 = k+1$. Por lo tanto:

$$(a_0 + a_1 + \cdots + a_{p-1})^2 = 1 + pk \equiv 1 \pmod{p}$$

luego $a_0 + a_1 + \cdots + a_{p-1} \equiv \pm 1 \pmod{p}$ o, equivalentemente, existe un entero t tal que

$$a_0 + a_1 + \cdots + a_{p-1} = \pm 1 + pt.$$

Ahora recordamos que los a_i son, por definición, los coeficientes del entero ciclotómico $\alpha/\bar{\alpha}$ en forma canónica, pero dichos coeficientes no son únicos, sino que podemos cambiar a_i por $a_i - t$ y seguimos teniendo unos coeficientes correspondientes a dicho entero, y con esta elección se cumple que $a_0 + \cdots + a_{p-1} = \pm 1$.

Todo lo dicho hasta aquí sigue siendo válido con esta elección concreta de los coeficientes, pero así llegamos a que $(a_0 + \cdots + a_{p-1})^2 = 1$, luego $k=0$, luego $b_0 = 1$, y (17.10) nos da entonces que todos los a_i son 0 excepto uno, que tiene que ser ± 1. Equivalentemente, hemos probado que

$$\frac{\alpha}{\bar{\alpha}} = E(\omega) = \pm\omega^r,$$

para cierto $r = 0, \ldots, p-1$, o también, $\alpha = \pm\omega^r\bar{\alpha}$.

Vamos a probar que el signo negativo no puede darse. Para ello consideramos el primo $\lambda = \omega - 1$. Hemos probado que todo entero ciclotómico es congruente con un entero racional módulo λ. Pongamos que $\alpha = k + \beta\lambda$, con lo que $\bar{\alpha} = k + \bar{\beta}\bar{\lambda} = k + \bar{\beta}\delta\lambda$, luego concluimos que $\alpha \equiv \bar{\alpha} \pmod{\lambda}$. Si fuera $\alpha = -\omega^r\bar{\alpha}$, como $\omega \equiv 1 \pmod{\lambda}$, sería $\alpha \equiv -\bar{\alpha} \pmod{\lambda}$, luego $\alpha \equiv -\alpha \pmod{\lambda}$, luego $\lambda \mid 2\alpha$, luego (como α es una unidad) $\lambda \mid 2$, pero tomando normas vemos que esto es imposible.

Así queda demostrado que $\alpha = \omega^r\bar{\alpha}$, para cierto exponente r. Pero los números $2\cdot 0, 2\cdot 1, \ldots, 2\cdot(p-1)$ recorren todas las clases de $\mathbb{Z}_p$, luego podemos tomar $r = 2m$, y así $\alpha = \omega^{2m}\bar{\alpha} = \omega^m\overline{\omega^{-m}}\bar{\alpha}$, o también $\overline{\alpha/\omega^m} = \alpha/\omega^m$. Esto significa que $\beta = \alpha/\omega^m$ es una unidad ciclotómica real y $\alpha = \omega^m\beta$. ∎

Con esto tenemos todos los hechos básicos de la aritmética de los enteros ciclotómicos que necesitaremos en la sección siguiente para demostrar el Último Teorema de Fermat para exponente $p = 5$. No obstante, dedicamos el resto de esta sección a mostrar que la aritmética de los enteros ciclotómicos de orden 5 se parece bastante, *mutatis mutandis* a la de los enteros cuadráticos.

Primos ciclotómicos Según el teorema 17.16 para encontrar primos ciclotómicos, sólo tenemos que factorizar primos racionales. La tabla siguiente contiene las descomposiciones en factores primos de los primos racionales menores que 100. Todos los factores son primos excepto el primer factor de la descomposición de 5, que es una unidad:

$$
\begin{aligned}
2 &= 2\\
3 &= 3\\
5 &= (\omega^3-\omega-1)(\omega-1)^4\\
7 &= 7\\
11 &= (\omega+2)(\omega^2+2)(\omega^3+2)(\omega^4+2)\\
13 &= 13\\
17 &= 17\\
19 &= (\omega^2+5\omega+1)(-5\omega^3-4\omega^2-4\omega-5)\\
23 &= 23\\
29 &= (\omega^2+6\omega+1)(-6\omega^3-5\omega^2-5\omega-6)\\
31 &= (\omega-2)(\omega^2-2)(\omega^3-2)(\omega^4-2)\\
37 &= 37\\
41 &= (-\omega^2+2\omega+1)(\omega^3+3\omega^2+\omega+2)(2\omega^3-\omega+1)(-3\omega^3-2\omega^2-2\omega-1)\\
43 &= 43\\
47 &= 47\\
53 &= 53\\
59 &= (2\omega^3-5\omega^2-5\omega+2)(7\omega^2+2\omega+7)\\
61 &= (-\omega^2+2\omega+2)(\omega^3+3\omega^2+\omega+3)(2\omega^3-\omega+2)(-3\omega^3-2\omega^2-2\omega)\\
67 &= 67\\
71 &= (2\omega^3-2\omega^2+2\omega+1)(2\omega^3+4\omega^2+4\omega+3)(-2\omega^2-4\omega-1)(-4\omega^3-2\omega-1)\\
73 &= 73\\
79 &= (3\omega^3-5\omega^2-5\omega+3)(8\omega^2+3\omega+8)\\
83 &= 83\\
89 &= (\omega^3-8\omega^2-8\omega+1)(9\omega^2+\omega+9)\\
97 &= 97
\end{aligned}
$$

Notemos que no es obvio cómo pueden obtenerse estas factorizaciones ni cómo justificar que los factores son primos. No obstante, aceptando la tabla podemos detectar en ella ciertos patrones. Por ejemplo, cada primo racional se descompone en producto de uno, dos o cuatro primos ciclotómicos, y es fácil reconocer el criterio que decide cuándo se da cada caso.

A continuación probamos el teorema análogo a 9.13 para enteros ciclotómicos. En el teorema siguiente, fijado un primo racional p, para cada $g(x) \in \mathbb{Z}[x]$, representamos por $\bar{g}(x)$ al polinomio de $\mathbb{Z}_p[x]$ que resulta de sustituir cada coeficiente de $g(x)$ por su resto módulo p.

Teorema 17.20 *Si p es primo y $\bar{c}_5(x) = \bar{g}_1(x)\cdots\bar{g}_r(x)$ es la descomposición del polinomio $\bar{c}_5(x)$ en factores primos en $\mathbb{Z}_p[x]$, entonces cada máximo común divisor $\pi_i = (p, g_i(\omega))$ es un primo ciclotómico, existe una unidad δ tal que $p = \delta\pi_1\cdots\pi_r$ y se cumple que π_i es asociado a π_j si y sólo si $\bar{g}_i(x)$ es asociado a $\bar{g}_j(x)$. Además, $\mathrm{N}(\pi_i) = p^f$, donde $f = \operatorname{grad} f_i(x)$.*

DEMOSTRACIÓN: Consideremos los ideales $\mathfrak{p}_i = (p, g_i(\omega)) = (\pi_i)$, donde la última igualdad viene dada por el teorema 13.5. Que π_i sea primo equivale a que lo sea el ideal $\mathfrak{p}_i$, y a su vez, en virtud de 13.13, basta probar que el anillo cociente $\mathbb{Z}[\omega]/\mathfrak{p}_i$ es un dominio íntegro. Consideremos el anillo cociente $\mathbb{Z}[x]/(p, g_i(x))$ y la aplicación

$$\mathbb{Z}[x]/(p, g_i(x)) \longrightarrow \mathbb{Z}[\omega]/(p, g_i(\omega))$$

dada por $[u(x)] \mapsto [u(\omega)]$.

Está bien definida, en el sentido de que $[u(x)] = [v(x)]$ es equivalente a que $u(x) - v(x) \in (p, g_i(x))$, es decir, a que existan polinomios $c(x)$ y $d(x)$ tales que $u(x) - v(x) = c(x)p + d(x)g_i(x)$, y esto implica que

$$u(\omega) - v(\omega) = c(\omega)p + d(\omega)g_i(\omega),$$

luego $u(\omega) - v(\omega) \in (p, g_i(\omega))$, luego $[u(\omega)] = [v(\omega)]$.

Recíprocamente, si dos clases tienen la misma imagen, es decir, si se cumple que $[u(\omega)] = [v(\omega)]$, entonces existen enteros ciclotómicos $c(\omega)$, $v(\omega)$ tales que

$$u(\omega) - v(\omega) = c(\omega)p + d(\omega)g_i(\omega),$$

luego existe un polinomio $h(x)$ (que surge de dividir euclídeamente un polinomio con coeficientes enteros entre $c_5(x)$, por lo que podemos asegurar que tiene coeficientes enteros) tal que

$$u(x) - v(x) = c(x)p + d(x)g_i(x) + h(x)c_5(x).$$

Ahora usamos que $\bar{g}_i(x)$ divide a $\bar{c}_5(x)$, es decir, que $\bar{c}_5(x) = \bar{g}_i(x)\bar{f}(x)$ o, equivalentemente, que existe un polinomio $q(x)$ tal que

$$c_5(x) = g_i(x)f(x) + pq(x).$$

Así

$$u(x) - v(x) = c(x)p + d(x)g_i(x) + h(x)f(x)g_i(x) + h(x)q(x)p \in (p, g_i(x)),$$

luego $[u(x)] = [v(x)]$.

Con esto hemos probado que cada clase de $\mathbb{Z}[x]/(p, g_i(x))$ se corresponde con una única clase de $\mathbb{Z}[\omega]/(p, g_i(\omega))$ y, obviamente, toda clase de este segundo anillo tiene una antiimagen. También es inmediato que esta correspondencia conserva sumas y productos, por lo que si en el segundo anillo hubiera dos clases no nulas cuyo producto fuera nulo, las clases correspondientes en el primer anillo cumplirían lo mismo, luego éste no sería un dominio íntegro. Por lo tanto, basta probar que $\mathbb{Z}[x]/(p, g_i(x))$ es un dominio íntegro.

Ahora usamos que, como $\bar{g}_i(x)$ es primo en $\mathbb{Z}_p[x]$, el ideal $(\bar{g}_i(x))$ es primo, luego el cociente $\mathbb{Z}_p[x]/(\bar{g}_i(x))$ es un dominio íntegro. Por lo tanto, basta probar que la aplicación

$$\mathbb{Z}[x]/(p, g_i(x)) \longrightarrow \mathbb{Z}_p[x]/(\bar{g}_i(x))$$

dada por $[u(x)] \mapsto [\bar{u}(x)]$ nos permite identificar ambos anillos igual que la anterior, pues entonces $\mathbb{Z}[x]/(p, g_i(x))$ será un dominio íntegro.

Nuevamente probamos que está bien definida, pues si $[u(x)] = [v(x)]$, entonces existen polinomios $c(x)$ y $d(x)$ tales que

$$u(x) - v(x) = c(x)p + d(x)g_i(x),$$

luego $\bar{u}(x) - \bar{v}(x) = \bar{d}(x)\bar{g}_i(x) \in (\bar{g}_i(x))$, luego $[\bar{u}(x)] = [\bar{v}(x)]$. Y, recíprocamente, si se cumple que $[\bar{u}(x)] = [\bar{v}(x)]$, entonces existe un polinomio $\bar{d}(x)$ tal que $\bar{u}(x) - \bar{v}(x) = \bar{d}(x)\bar{g}_i(x)$, lo que equivale a que exista un polinomio $c(x)$ tal que

$$u(x) - v(x) = c(x)p + d(x)g_i(x) \in (p, g_i(x)),$$

con lo que $[u(x)] = [v(x)]$. Esto prueba que las clases de uno de los anillos se corresponden biunívocamente con las del otro y, como la correspondencia conserva sumas y productos, es inmediato que uno es un dominio íntegro si y sólo si lo es el otro.

Supongamos que π_i y π_j son asociados. Entonces $\pi_i \mid g_i(\omega)$ y $\pi_i \mid g_j(\omega)$.

Dos enteros racionales a y b son congruentes módulo π_i si y sólo si lo son módulo p, pues una implicación es obvia y, si $a \equiv b \pmod{\pi}$, entonces $\pi \mid a - b$, luego $p \mid \mathrm{N}(\pi) \mid \mathrm{N}(a - b) = (a - b)^4$, luego $p \mid a - b$, luego $a \equiv b \pmod{p}$.

Esto se traduce en que las clases con representantes enteros racionales en el anillo de clases de restos $k = \mathbb{Z}[\omega]_{\pi_i}$ forman un cuerpo que podemos identificar con $\mathbb{Z}_p$. Notemos que k es finito, pues toda clase admite un representante con coeficientes a_0, a_1, a_2, a_3 menores que p (ya que $p \equiv 0 \pmod{\pi_i}$), luego hay a lo sumo p^4 clases de restos módulo π_i. Como π_i es primo el cociente es un dominio íntegro finito, luego es un cuerpo finito (teorema 13.14).

Tenemos que $\bar{\omega} \in k$ es una raíz del polinomio $\bar{g}_i(x)$, y también de $\bar{g}_j(x)$. Sea $\bar{f}(x) \in \mathbb{Z}_p[x]$ un polinomio del menor grado posible que tenga a $\bar{\omega}$ por raíz en k. Dividimos

$$\bar{g}_i(x) = \bar{f}(x)\bar{q}_i(x) + \bar{r}_i(x), \qquad \bar{g}_j(x) = \bar{f}(x)\bar{q}_j(x) + \bar{r}_j(x)$$

con restos de grado menor que el del divisor, y al evaluar en $\bar{\omega}$ concluimos que $\bar{r}_i(\bar{\omega}) = \bar{r}_j(\bar{\omega}) = 0$, luego por la minimalidad del grado, $\bar{r}_i(x) = \bar{r}_j(x) = 0$, luego $\bar{f}(x) \mid \bar{g}_i(x)$ y $\bar{f}(x) \mid \bar{g}_j(x)$, pero $\bar{g}_i(x)$ y $\bar{g}_j(x)$ son irreducibles, luego tiene que ser $\bar{g}_i(x) = \bar{c}\bar{g}_j(x)$, para cierto $\bar{c} \in \mathbb{Z}_p$ no nulo, y concluimos que $\bar{g}_i(x)$ y $\bar{g}_j(x)$ son asociados.

Recíprocamente, si $\bar{g}_i(x) = \bar{c}\bar{g}_j(x)$, con $\bar{c} \in \mathbb{Z}_p$ no nulo, entonces

$$g_i(x) = cg_j(x) + ph(x),$$

para cierto polinomio h, y entonces $g_i(\omega) = cg_j(\omega) + ph(\omega)$, de donde se sigue que $\pi_j \mid g_i(\omega)$, luego $\pi_j \mid \pi_i$ y, como ambos son primos, de hecho son asociados.

Ahora distinguimos casos según las posibles factorizaciones de $\bar{c}_5(x)$.

1. $\bar{c}_5(x)$ es irreducible en $\mathbb{Z}_p$.

   Entonces $r = 1$ y $g_1(x) = c_5(x)$, luego, como $g_1(\omega) = 0$, resulta que $\pi_1 = (p, c_5(\omega)) = p$, luego concluimos que π_1 es un primo ciclotómico de norma p^4, donde $4 = \operatorname{grad} \bar{g}_1(x)$.

2. $\bar{c}_5(x)$ tiene un factor de grado 1, digamos $\bar{g}_1(x) = x - \bar{c}$.

   Entonces $\pi_1 = (p, \omega - c)$ es un divisor primo de p tal que $\omega \equiv c$ (mód π_1). Si, como antes, llamamos $k = \mathbb{Z}[\omega]_\pi$, tenemos que las raíces de $\bar{c}_5(x)$ que son las potencias $\bar{\omega}^i = \bar{c}^i$, están en $\mathbb{Z}_p$, luego

   $$\bar{c}_5(x) = (x - \bar{c})(x - \bar{c}^2)(x - \bar{c}^3)(x - \bar{c}^4).$$

   Si $p \neq 5$, entonces $c_5(\bar{1}) = \bar{5} \neq 0$, luego $\bar{c} \neq \bar{1}$, luego las cuatro potencias $\bar{c}^i$ son distintas dos a dos, ya que el orden de la clase $\bar{c}$ en el grupo k^* de las unidades de k divide a 5 y no es 1, luego es 5.

   Por lo tanto, en este caso todos los factores irreducibles de $\bar{c}_5(x)$ son de grado 1 y son no asociados dos a dos, luego tenemos cuatro primos no asociados $\pi_1, \pi_2, \pi_3, \pi_4$ que dividen a p, luego su producto también divide a p y, como p no puede tener más de cuatro factores primos, concluimos que $p = \delta\pi_1\pi_2\pi_3\pi_4$, como afirma el enunciado. Además $\mathrm{N}(\pi_i) = p^1$, donde el exponente 1 es el grado de cada $\bar{g}_i(x)$.

   Si $p = 5$ tenemos, por una parte, que $x^5 - \bar{1} = (x - \bar{1})^5$, de donde concluimos que $\bar{c}_5(x) = (x - \bar{1})^4$, luego $g_i(x) = x - 1$ para todo i, luego $\pi_i = \omega - 1 = \lambda$ y, por otro lado, sabemos que $5 = \delta\lambda^4$, con $\mathrm{N}(\lambda) = p^1$, luego también se cumple el teorema en este caso.

   Además, puesto que hemos visto que si un factor de $\bar{c}_5(x)$ tiene grado 1 lo mismo les sucede a los demás, sólo hay una forma más de factorización posible:

3. $\bar{c}_5(x) = \bar{g}_1(x)\bar{g}_2(x)$, con $\bar{g}_1(x)$, $\bar{g}_2(x)$ irreducibles en $\mathbb{Z}_p$ (de grado 2).

   Entonces no puede ser $p = 5$, pues ya sabemos que entonces se da el caso precedente. Tenemos dos primos π_1, π_2 que dividen a p. Vamos a ver que no pueden ser asociados.

   Consideramos el cuerpo k_{π_1}, en el que el polinomio $\bar{c}_5(x)$ tiene cuatro raíces distintas dos a dos, pues $\bar{\omega} \neq \bar{1}$ (ya que $c_5(\bar{1}) \neq \bar{0}$), luego, al igual que antes, concluimos que $\bar{\omega}$ tiene orden 5 en el grupo k^*. Así pues, el polinomio $\bar{c}_5(x)$ tiene cuatro raíces distintas en k, pero si $\bar{g}_1(x)$ y $\bar{g}_2(x)$ fueran asociados en $\mathbb{Z}_p[x]$, entonces tendrían las mismas raíces en k, luego su producto $\bar{c}_5(x)$ tendría raíces repetidas.

   Basta probar que $p = \delta\pi_1\pi_2$, donde δ es una unidad, pues entonces tiene que ser $\mathrm{N}(\pi_i) = p^2$ y el exponente es el grado de los polinomios $\bar{g}_i(x)$. Equivalentemente, hay que probar que p es asociado a $\pi_1\pi_2$. Ciertamente,

$\pi_1\pi_2 \mid p$, pues π_1, π_2 son dos divisores primos de p no asociados, luego su producto divide a p. Basta probar que $p \mid \pi_1\pi_2$.

Por la relación de Bezout existen enteros ciclotómicos tales que

$$\pi_1 = \alpha_1 p + \beta_1 g_1(\omega), \qquad \pi_2 = \alpha_2 p + \beta_2 g_2(\omega),$$

luego

$$\pi_1\pi_2 = \alpha_1\alpha_2 p^2 + \alpha_2\beta_1 g_1(\omega)p + \alpha_1\beta_2 g_2(\omega)p + \beta_1\beta_2 g_1(\omega)g_2(\omega),$$

y basta tener en cuenta que $g_1(x)g_2(x) \equiv c_5(x)$ (mód p), luego

$$g_1(x)g_2(x) = c_5(x) + ph(x),$$

para cierto polinomio $h(x)$ y, evaluando en ω, resulta que

$$g_1(\omega)g_2(\omega) = ph(\omega),$$

con lo que llegamos a que $p \mid \pi_1\pi_2$. ∎

En la prueba del teorema anterior hemos obtenido un poco más de lo que afirma el enunciado:

Teorema 17.21 *Si p es un primo racional, entonces su descomposición en factores primos ciclotómicos tiene que ser de uno de los tipos siguientes:*

1. *$p = \delta\pi^4$ (lo cual sólo sucede para $p = 5$ con $\pi = \lambda$).*
2. *$p = \pi_1\pi_2\pi_3\pi_4$, donde los factores tienen todos norma p y son no asociados dos a dos.*
3. *$p = \pi_1\pi_2$, donde los factores tienen norma p^2 y no son asociados.*
4. *p es un primo ciclotómico (de norma p^4).*

Ejemplo Ahora ya sabemos cómo obtener las factorizaciones que hemos mostrado en la tabla precedente. Consideremos, por ejemplo, el caso de $p = 19$. Se cumple que

$$x^4 + x^3 + x^2 + x + 1 \equiv \left(x^2 + 5x + 1\right)\left(x^2 + 15x + 1\right) \text{ (mód 19)}$$

(lo cual puede comprobarse mediante un proceso finito, ya que el número de factores posibles es finito). Por lo tanto, 19 tiene dos factores primos de norma 19^2 y que son, concretamente:

$$\pi_1 = (19, \omega^2 + 5\omega + 1), \qquad \pi_2 = (19, \omega^2 + 15\omega + 1).$$

Pero $\mathrm{N}(\omega^2 + 5\omega + 1) = 19^2$, luego $\pi_1 = \omega^2 + 5\omega + 1$ ya es primo, y π_2 puede obtenerse como

$$\frac{19}{\omega^2 + 5\omega + 1} = -5\omega^3 - 4\omega^2 - 4\omega - 5.$$

∎

Sólo falta probar un hecho que se aprecia claramente en la tabla de factorizaciones: el tipo de factorización de un primo p depende de su resto módulo 5. Para probarlo demostramos primero lo siguiente:

Teorema 17.22 *Si π es un primo ciclotómico, entonces el número de clases de restos módulo π es* $\mathrm{N}(\pi)$.

DEMOSTRACIÓN: Pongamos que $\mathrm{N}(\pi) = p^f$ y consideremos el cuerpo de clases de restos $K = \mathbb{Z}[\omega]_\pi$, que contiene a $\mathbb{Z}_p$. Basta probar que cada elemento de K se expresa de forma única como polinomio en $\bar\omega$ de grado menor que f con coeficientes en $\mathbb{Z}_p$, pues entonces el número de elementos de K será p^f (pues cada polinomio está determinado por f coeficientes que pueden tomar p valores posibles cada uno).

Según el teorema 17.20, la descomposición de $c_5(x)$ en factores primos módulo p es de la forma

$$\bar c_5(x) = \bar g_1(x)\cdots\bar g_r(x),$$

donde cada polinomio $g_i(x)$ tiene grado f. Entonces, en K se cumple que

$$\bar g_1(\bar\omega)\cdots\bar g_r(\bar\omega) = \bar c_5(\bar\omega) = 0,$$

luego existe un i tal que $\bar g_i(\bar\omega) = 0$. Sea $\bar h(x) \in \mathbb{Z}_p[x]$ un polinomio no nulo del menor grado posible que tenga a $\bar\omega$ por raíz. Dividimos $\bar g_i(x) = \bar h(x)\bar q(x) + \bar r(x)$, donde $\bar r(x)$ es nulo o su grado es menor que el de $\bar h$. Pero evaluando en $\bar\omega$ queda que $\bar r(\bar\omega) = 0$, luego por la minimalidad del grado de $\bar h$ tiene que ser $\bar r(x) = 0$, luego $\bar h(x) \mid \bar g_i(x)$, pero $\bar g_i(x)$ es irreducible, luego $\bar g_i(x) = \bar h(x)$ y con esto hemos probado que f es menor grado posible de un polinomio no nulo que tenga a $\bar\omega$ entre sus raíces.

Si $h(\omega)$ es un entero ciclotómico arbitrario, dividimos $h(x) = g_i(x)q(x)+r(x)$, donde r es nulo o tiene grado menor que f. Entonces $\bar h(\bar\omega) = \bar r(\bar\omega)$, luego, en efecto, todo elemento de K se expresa como polinomio en $\bar\omega$ de grado menor que f. Falta probar que la expresión es única, pero si $\bar r_1(\bar\omega) = \bar r_2(\bar\omega)$, donde $r_1(x)$ y $r_2(x)$ tienen grado menor que f, entonces $\bar r_1(x) - \bar r_2(x)$ tiene también grado menor que f y se anula en $\bar\omega$, luego, según hemos visto, tiene que ser el polinomio nulo, luego $\bar r_1(x) = \bar r_2(x)$, y esto significa que las dos expresiones de las que hemos partido son en realidad la misma. ∎

Ahora ya podemos determinar el tipo de factorización de un primo en función de su resto módulo 5:

Teorema 17.23 *Si π es un primo ciclotómico y $p \neq 5$ es el primo racional al cual divide, entonces* $\mathrm{N}(\pi) = p^f$*, donde $f = o_5(p)$ es el orden de p en U_5.*

DEMOSTRACIÓN: Sea $\mathrm{N}(\pi) = p^f$. Tenemos que demostrar que $f = o_5(p)$. Por el teorema anterior el cuerpo $K = \mathbb{Z}[\omega]_\pi$ tiene p^f elementos, luego su grupo de unidades tiene orden $p^f - 1$. En dicho cuerpo la clase $\bar\omega$ tiene orden 5 (notemos que no puede ser $\bar\omega = 1$, pues entonces $\pi \mid \omega - 1 \mid 5$, y sería $p = 5$). Por consiguiente, $5 \mid p^f - 1$, luego $p^f \equiv 1$ (mód 5), luego $o_5(p) \mid f$.

Por otra parte, un elemento cualquiera de K es de la forma

$$a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0,$$

luego un elemento cualquiera de K es de la forma

$$\alpha = \bar{a}_3\bar{\omega}^3 + a_2\bar{\omega}^2 + a_1\bar{\omega} + \bar{a}_0.$$

Como K tiene característica p, si llamamos $e = o_5(p)$, tenemos que

$$\alpha^{p^e} = (\bar{a}_3)^{p^e}(\bar{\omega}^{p^e})^3 + (a_2)^{p^e}(\bar{\omega}^{p^e})^2 + (a_1)^{p^e}(\bar{\omega}^{p^e}) + (\bar{a}_0)^{p^e}.$$

Ahora bien, por una parte, todo número entero cumple $a^p \equiv a$ (mód p) (luego también módulo π) y por otra $p^e \equiv 1$ (mód 5), luego $\omega^{p^e} = \omega$, y así, $\alpha^{p^e} = \alpha$. Esto significa que los p^f elementos de K son raíces del polinomio $x^{p^e} - x$, lo que implica que $p^f \le p^e$, luego $f \le e$ y, de hecho, se da la igualdad $f = e$. ∎

Explícitamente, el valor de f en función del resto de p módulo 5 es:

p (mód 5)	1	2	3	4
f	1	4	4	2

y con esto queda probado el comportamiento que mostraba la tabla de factorizaciones: los primos que se descomponen en cuatro factores primos ciclotómicos son (aparte del 5) los que cumplen $p \equiv 1$ (mód 5), mientras que los que se descomponen en dos factores son los que cumplen $p \equiv 2, 3$ (mód 5) y los que se conservan primos son los que cumplen $p \equiv -1$ (mód 5).

17.7 El Último Teorema de Fermat para p = 5

Pasamos ya a demostrar el Último Teorema de Fermat para $p = 5$. Más en general, probaremos que la ecuación

$$x^p + y^p = z^p$$

no tiene soluciones enteras no triviales (en las que ninguna variable valga 0) para todo primo p que cumpla dos hipótesis: una es que el anillo $\mathbb{Z}[\omega]$ de enteros ciclotómicos tenga factorización única, cosa que ya hemos demostrado para el exponente $p = 5$, y la segunda la demostraremos a continuación.

En primer lugar probamos un hecho elemental:

Teorema 17.24 *Un entero ciclotómico $\alpha = a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0$ en forma canónica es congruente con un entero racional módulo p si y sólo si $a_{p-1} \equiv a_{p-2} \equiv \cdots \equiv a_2 \equiv a_1$ (mód p).*

Demostración: Supongamos en primer lugar que $a_{p-1} = 0$, es decir, que α está en forma reducida. Entonces α es congruente con un entero racional k módulo p si y sólo si

$$\alpha = a_{p-2}\omega^{p-2} + \cdots + a_1\omega + a_0 = k + p(b_{p-2}\omega^{p-2} + \cdots + b_1\omega + b_0),$$

para ciertos enteros racionales $b_0, \ldots, b_{p-2}$. Por la unicidad de la forma reducida, esto equivale a que $a_i = 5b_i$, para $i = 1, \ldots, p-2$ y $a_0 = k + pb_0$. Claramente, esto sucede si y sólo si $p \mid a_i$, para $i = 1, \ldots, p-2$.

En el caso general aplicamos este criterio a la forma reducida

$$\alpha = (a_{p-2} - a_{p-1})\omega^{p-2} + \cdots + (a_1 - a_{p-1})\omega + a_0 - a_{p-1},$$

de modo que α es congruente con un entero módulo p si y sólo si $p \mid a_i - a_{p-1}$, para $i = 1, \ldots, p-2$, lo cual equivale la condición del enunciado. ∎

Otro hecho obvio es el siguiente:

Teorema 17.25 *Toda potencia p-esima de un entero ciclotómico es congruente con un entero módulo p.*

DEMOSTRACIÓN: Si α es una potencia p-ésima, entonces

$$\alpha = (a_{p-2}\omega^{p-2} + \cdots + a_1\omega + a_0)^p \equiv a_{p-2}^p\omega^{p(p-2)} + \cdots + a_1^p\omega^p + a_0^p$$
$$= a_{p-2}^p + \cdots + a_1^p + a_0^p \text{ (mód } p).$$

∎

Lo que no es obvio en absoluto (y no es cierto para p arbitrario) es que las unidades ciclotómicas quintas cumplen el recíproco:

Teorema 17.26 (Lema de Kummer 2) *Si α es una unidad ciclotómica de orden 5, entonces α es una potencia quinta si y sólo si es congruente con un entero módulo 5.*

DEMOSTRACIÓN: Una implicación es el teorema anterior. Supongamos que α es una unidad y es congruente con un entero k módulo 5. Pongamos que $\alpha = \pm\omega^m\epsilon^n$. Podemos suponer que el signo es positivo, pues en caso contrario $-\alpha$ se expresa de esta forma con signo positivo, y si $-\alpha$ es congruente con un entero, α también.

Sea $n = 5c + r$, con $0 \le r < 5$. Entonces $\alpha = \omega^m(\epsilon^c)^5\epsilon^r$. Por la implicación ya probada sabemos que $(\epsilon^c)^5 \equiv k'$ (mód 5), para un cierto entero k' que no puede ser 0, pues entonces el miembro izquierdo sería múltiplo de 5, cuando es una unidad. Por lo tanto, existe un entero k'' tal que $k'k'' \equiv 1$ (mód 5), luego

$$\omega^m\epsilon^r \equiv k'k''\omega^m\epsilon^r \equiv k''\omega^m(\epsilon^c)^5\epsilon^r \equiv k''k \text{ (mód } 5),$$

luego $\omega^m\epsilon^r$ también es congruente con un entero módulo 5. Si probamos que es una potencia quinta, lo mismo valdrá al multiplicar por $(\epsilon^c)^5$, es decir, lo mismo valdrá para α.

Ahora calculamos

$$\epsilon^0 = 1, \quad \epsilon = -\omega^3 - \omega^2, \quad \epsilon^2 = -\omega^3 - \omega^2 + 1,$$
$$\epsilon^3 = -2\omega^3 - 2\omega^2 + 1, \quad \epsilon^4 = -3\omega^3 - 3\omega^2 + 2$$

y vemos que, según el criterio dado por el teorema anterior, la única potencia congruente con un entero módulo 5 es $\epsilon^0 = 1$. El efecto de multiplicar estas potencias por ω^m es que se permutan sus coeficientes, luego para que un producto $\omega^m\epsilon^r$ sea congruente con un entero racional módulo 5 es necesario que cuatro

de los coeficientes de ϵ^r sean congruentes módulo 5, y esto no sucede salvo para $r = 0$, en cuyo caso los coeficientes en cuestión son a_1, a_2, a_3, a_4, por lo que necesariamente $m = 0$, luego $\omega^m\epsilon^r = 1$ es trivialmente una potencia quinta. ∎

Ahora ya podemos probar:

Teorema 17.27 *Sea $p \geq 5$ un primo que cumpla las dos condiciones siguientes:*

1. *El anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos de orden p tiene factorización única.*
2. *Una unidad ciclotómica de orden p es una potencia p-ésima si y sólo si es congruente con un entero módulo p.*

Entonces la ecuación $x^p + y^p = z^p$ no tiene soluciones enteras no triviales. En particular, esto vale para $p = 5$.

Demostración: Supongamos que existen números enteros no nulos x, y, z tales que $x^p + y^p = z^p$. Podemos suponer que son primos entre sí dos a dos, pues si un primo q divide a dos de ellos, la ecuación implica que también divide al otro, y entonces, $x/q, y/q, z/q$ también cumple la ecuación. Procediendo de este modo, tras un número finito de pasos llegamos a una solución en la que los tres números son primos entre sí dos a dos.

En particular, p divide a lo sumo a uno de los tres, y en tal caso no perdemos generalidad si suponemos que divide a z, pues si, por ejemplo, divide a x, la ecuación se puede poner en la forma

$$y^p + (-z)^p = (-x)^p,$$

y así p divide a la variable del miembro derecho. Por lo tanto, no perdemos generalidad si consideramos únicamente los dos casos siguientes:

Caso I p no divide a ninguno de los números x, y, z.

Caso II p divide a z (pero no a x ni a y).

Consideramos primero el Caso I. Empezamos factorizando:

$$x^p + y^p = (x + y)(x + \omega y)(x + \omega^2 y)\cdots(x + \omega^{p-1}y) \tag{17.11}$$

y vamos a probar que los p factores son primos entre sí dos a dos.

Supongamos, por el contrario, que dos de ellos, $x + \omega^i y$, $x + \omega^j y$, con $i < j$ tienen un divisor primo común $\mathfrak{p}$. Entonces

$$\mathfrak{p} \mid (x+\omega^j y)-(x+\omega^i y) = \omega^i(\omega^{j-i}-1)y,\ \mathfrak{p} \mid \omega^{j-i}(x+\omega^i y)-(x+\omega^j y) = (\omega^{j-i}-1)x,$$

pero $\mathfrak{p}$ no puede dividir a $\omega^{j-i} - 1$, pues entonces sería asociado del primo $\lambda = \omega - 1$, luego λ dividiría al miembro derecho de (17.11), es decir, a z^p, luego a z, luego tomando normas $p \mid z$, en contra de lo supuesto.

Por lo tanto, $\mathfrak{p}$ divide a x y a y, pero si $\mathrm{N}(\mathfrak{p}) = q^f$, entonces $q \mid x$ y $q \mid y$, en contra de que x e y son primos entre sí.

Ahora usamos una de las ideas fundamentales de la demostración (que se basa en que $\mathbb{Z}[\omega]$ tiene factorización única): como el miembro derecho de (17.11) es una potencia p-ésima y los factores son primos entre sí, la multiplicidad de cada primo en cada factor tiene que ser múltiplo de p, lo cual se traduce en que cada factor es una potencia quinta salvo una unidad. En particular:

$$x + \omega y = \alpha\beta^p,$$

donde α es una unidad ciclotómica. Aplicando la conjugación compleja:

$$x + \omega^{-1} y = \bar{\alpha}\bar{\beta}^p.$$

Ahora usamos el lema de Kummer 17.19, que nos asegura que $\alpha = \omega^m\delta$, donde δ es una unidad ciclotómica real. Por lo tanto $\bar{\alpha} = \omega^{-m}\delta$ y

$$\frac{\alpha}{\bar{\alpha}} = \omega^{2m} = \omega^r,$$

para $0 \le r < 5$.

Por otra parte, por 17.25 toda potencia quinta es congruente con un entero módulo 5, y el argumento vale igualmente para todo p, luego existen un entero racional k y un entero ciclotómico θ tales que

$$\beta^p = k + p\theta, \quad \text{luego} \quad \bar{\beta}^p = k + p\bar{\theta}$$

y concluimos que $\beta^p \equiv k \equiv \bar{\beta}^p \pmod{p}$. Juntando todo esto obtenemos que

$$x + \omega^{-1} y = \bar{\alpha}\bar{\beta}^p = \omega^{-r}\alpha\bar{\beta}^p \equiv \omega^{-r}\alpha\beta^p = \omega^{-r}(x + \omega y) \pmod{p}.$$

Equivalentemente:

$$p \mid x\omega^r + y\omega^{r-1} - y\omega - x.$$

Ahora observamos que para que un entero ciclotómico en forma reducida sea múltiplo de p es necesario que todos sus coeficientes sean múltiplos de p, y si está en forma canónica, es necesario que todos sus coeficientes sean congruentes módulo p. Si en la fórmula precedente $r > 2$, entonces todas las potencias de ω son distintas entre sí y, como $p \ge 5$ y sólo hay 4 términos, una de las potencias de ω tiene coeficiente 0, luego todos los coeficientes tienen que ser múltiplos de p, luego $p \mid x$ y $p \mid y$, en contra de lo supuesto.

Si $r = 2$ tenemos que $p \mid x\omega^2 - x$, y el mismo razonamiento nos lleva a que $p \mid x$ y de nuevo tenemos una contradicción. Lo mismo ocurre si $r = 0$, pues entonces tenemos que $p \mid y\omega^{p-1} - y\omega$. Concluimos que tiene que ser $r = 1$, en cuyo caso lo que tenemos es que $p \mid (x - y)\omega - (x - y)$, de donde únicamente deducimos que $x \equiv y \pmod{p}$.

Así pues, bajo la hipótesis del caso I, hemos demostrado que $x \equiv y \pmod{p}$, pero el caso I es simétrico. Como la terna $(x, -z, -y)$ también es una solución de la ecuación de Fermat que cumple el caso I, también podemos concluir que $x \equiv -z \pmod{p}$, luego

$$0 = x^p + y^p + (-z)^p \equiv 3x^p \pmod{p},$$

y nuevamente llegamos a que $p \mid x$, y esta contradicción termina la prueba del caso I.

Abordamos ahora el caso II. Pongamos que $z = p^k z'$, donde $p \nmid z'$. La ecuación de Fermat equivale ahora a que

$$x^p + y^p = p^{pk} z'^p,$$

donde p no divide a x, y, z'. Usando que $p = \delta_0 \lambda^{p-1}$, para cierta unidad δ_0, tenemos que

$$x^p + y^p = (\delta_0)^{pk} \lambda^{p(p-1)k} z'^p = \delta \lambda^{pm} z'^p,$$

para cierta unidad ciclotómica δ y cierto $m > 0$, donde x, y, z' no son divisibles entre p.

Vamos a llegar a una contradicción demostrando algo más general, a saber, que la ecuación

$$\xi^p + \upsilon^p = \delta \lambda^{pm} \zeta^p \tag{17.12}$$

no admite soluciones $\xi, \upsilon, \zeta, \delta$ que sean enteros ciclotómicos, tales que δ sea una unidad y que ξ, υ, ζ no sean divisibles entre λ.

Si existe una solución en estas condiciones, podemos tomar una en la que el exponente $m > 0$ tome el menor valor posible. Vamos a probar que, necesariamente, $m > 1$ y que existe otra solución con $m - 1$ en lugar de m, lo que contradice a la minimalidad de m.

Como en el caso anterior, podemos factorizar:

$$\delta \lambda^{pm} \zeta^p = \xi^p + \upsilon^p = (\xi + \upsilon)(\xi + \omega\upsilon) \cdots (\xi + \omega^{p-1}\upsilon), \tag{17.13}$$

pero ahora no podemos asegurar que los factores de la derecha sean primos entre sí. De hecho, ni siquiera estamos suponiendo que ξ, υ, ζ sean primos entre sí dos a dos. Lo que podemos decir de los factores es lo siguiente:

> *λ divide a todos los factores $\xi + \omega^i \upsilon$, luego en particular divide a sus diferencias, pero si $i < j$, entonces $\lambda^2 \nmid (\xi + \omega^j \upsilon) - (\xi + \omega^i \upsilon)$.*

En efecto, como en el caso I, tenemos que

$$(\xi + \omega^j \upsilon) - (\xi + \omega^i \upsilon) = \omega^i(\omega^{j-i} - 1)\upsilon = \delta_0 \lambda \upsilon,$$

para cierta unidad ciclotómica δ_0. Si $\lambda^2 \mid (\xi + \omega^j \upsilon) - (\xi + \omega^i \upsilon)$, entonces $\lambda^2 \mid \delta_0 \lambda \upsilon$, luego $\lambda \mid \upsilon$, en contradicción con lo supuesto.

Por otra parte, tenemos que λ divide al miembro izquierdo de (17.13), luego divide a uno de los factores de la derecha, pero como también divide a la diferencia de dos cualesquiera de ellos, tiene que dividirlos a todos.

Por lo tanto, los enteros ciclotómicos $(\xi + \omega^i \upsilon)/\lambda$, para $i = 0, \ldots, p-1$, son no congruentes módulo λ dos a dos, pues si λ dividiera a la diferencia de dos de ellos, entonces λ^2 dividiría a la diferencia de los numeradores, y acabamos de probar que no es así.

Pero sólo hay p clases de congruencia módulo λ, luego uno de los p enteros ciclotómicos anteriores tiene que ser congruente con 0 módulo λ, es decir, que $\lambda^2 \mid \xi + \omega^i \upsilon$, para un cierto i.

En resumen: hemos probado que λ divide con multiplicidad 1 a cada uno de los factores del miembro derecho de (17.13) excepto a uno de ellos, que es divisible al menos con multiplicidad 2.

Observamos ahora que no perdemos generalidad si suponemos que, concretamente $i = 0$, es decir, que $\lambda^2 \mid \xi + \upsilon$.

En efecto, en principio tenemos que $\lambda^2 \mid \xi + \omega^i \upsilon$, pero podemos llamar υ a $\omega^i \upsilon$ y, como $(\omega^i \upsilon)^p = \upsilon^p$, se sigue cumpliendo la ecuación (17.13) y el nuevo υ sigue sin ser divisible entre λ (ya que sólo hemos multiplicado el viejo por una unidad), pero ahora es $\xi + \upsilon$ el factor divisible entre λ^2.

Recapitulando: λ divide exactamente una vez a cada factor del miembro derecho de (17.13) excepto al primero, al cual divide al menos dos veces, luego en total divide al miembro derecho con multiplicidad al menos $p + 1$ veces, lo cual obliga a que en el miembro izquierdo sea $m > 1$, como queríamos probar. Más aún, podemos precisar que λ divide a $\xi + \upsilon$ con multiplicidad exactamente $p(m-1)+1$, es decir, las pm veces que divide al miembro izquierdo menos $p-1$, pues los otros $p - 1$ factores son divisibles una vez cada uno.

Llamemos $\mathfrak{m}$ al máximo común divisor de ξ, υ. Como λ no divide a ninguno de los dos, también $\lambda \nmid \mathfrak{m}$. Para $i \neq 0$ tenemos que

$$\xi + \omega^i \upsilon = (\lambda)\mathfrak{m}\mathfrak{c}_i,$$

para cierto entero ciclotómico $\mathfrak{c}_i$ no divisible entre λ (porque, como $\mathfrak{m}$ divide a ξ y a υ, también divide a $\xi + \omega^i \upsilon$, que además sabemos que es divisible entre λ exactamente una vez). Para $i = 0$ cambia el exponente de λ:

$$\xi + \upsilon = (\lambda)^{p(m-1)+1}\mathfrak{m}\mathfrak{c}_0.$$

Veamos ahora que los enteros $\mathfrak{c}_i$ son primos entre sí dos a dos.

En efecto, si un primo ciclotómico $\mathfrak{p}$ (que no puede ser λ) divide a $\mathfrak{c}_i$ y $\mathfrak{c}_j$, con $i < j$, entonces $\mathfrak{m}\mathfrak{p}$ divide a $\xi + \omega^i \upsilon$ y a $\xi + \omega^j \upsilon$, luego

$$\mathfrak{m}\mathfrak{p} \mid (\xi + \omega^j \upsilon) - (\xi + \omega^i \upsilon) = \omega^i(\omega^{j-i} - 1)\upsilon,$$

$$\mathfrak{m}\mathfrak{p} \mid \omega^{j-i}(\xi + \omega^i \upsilon) - (\xi + \omega^j \upsilon) = (\omega^{j-i} - 1)\xi.$$

Como $\omega^{j-i} - 1$ es un primo asociado a λ, que no divide a $\mathfrak{m}\mathfrak{p}$, tiene que ser $\mathfrak{m}\mathfrak{p} \mid \xi$, $\mathfrak{m}\mathfrak{p} \mid \upsilon$, pero esto contradice que $\mathfrak{m}$ sea el máximo común divisor de ξ, υ.

La ecuación (17.13) se convierte ahora en

$$\mathfrak{m}^p(\lambda)^{pm}\mathfrak{c}_0 \cdots \mathfrak{c}_{p-1} = (\delta)(\lambda)^{pm}(\zeta)^p.$$

Equivalentemente,

$$\mathfrak{c}_0 \cdots \mathfrak{c}_{p-1} = (\delta)\left(\frac{(\zeta)}{\mathfrak{m}}\right)^p,$$

donde el cociente es un entero ciclotómico.

Ahora, como los factores de la izquierda son primos entre sí dos a dos y el miembro derecho es una potencia p-ésima (salvo una unidad, pero eso aquí no importa), cada primo $\mathfrak{p}$ que divide a un $\mathfrak{c}_i$ tiene que hacerlo con multiplicidad divisible entre p, luego $\mathfrak{c}_i$ es una potencia p-ésima salvo una unidad, es decir, $\mathfrak{c}_i = (\eta_i)\mathfrak{b}_i^p$, donde η_i es una unidad y $\lambda \nmid \mathfrak{b}_i$. Por lo tanto:

$$\xi + \upsilon = (\eta_0)(\lambda)^{p(m-1)+1}\mathfrak{m}\mathfrak{b}_0^p, \qquad \xi + \omega^i\upsilon = (\eta_i)(\lambda)\mathfrak{m}\mathfrak{b}_i^p, \quad i = 1, \dots, p-1.$$

Despejamos $\mathfrak{m}$ en la primera ecuación y lo sustituimos en las otras:

$$(\eta_0)(\lambda)^{p(m-1)}\mathfrak{b}_0^p(\xi + \omega^i\upsilon) = (\eta_i)(\xi + \upsilon)\mathfrak{b}_i^p, \qquad i = 1, \dots, p-1.$$

Todo entero ciclotómico divide a su norma, por lo que $\bar{\mathfrak{b}}_0 = N(\mathfrak{b}_0)/\mathfrak{b}_0$ es un entero ciclotómico no divisible entre λ, porque en caso contrario $\lambda \mid \mathrm{N}(\mathfrak{b}_0)$, luego λ divide a alguno de los conjugados de $\mathfrak{b}_0$ (cuyo producto es la norma), pero, como λ es asociado de sus conjugados, llegamos a que $\lambda \mid \mathfrak{b}_0$, lo cual es falso. Multiplicando por $\mathfrak{b}_0^p$ queda:

$$(\eta_0)(\lambda)^{p(m-1)}N(\mathfrak{b}_0)^p(\xi + \omega^i\upsilon) = (\eta_i)(\xi + \upsilon)(\bar{\mathfrak{b}}_0\mathfrak{b}_i)^p, \qquad i = 1, \dots, p-1.$$

Llamando $(\alpha_i) = \bar{\mathfrak{b}}_0\mathfrak{b}_i$ (que es un entero ciclotómico no divisible entre λ) y agrupando las unidades, queda:

$$\lambda^{p(m-1)}N(\mathfrak{b}_0)^p(\xi + \omega^i\upsilon) = \delta_i(\xi + \upsilon)\alpha_i^p, \qquad i = 1, \dots, p-1.$$

Vamos a expresar estas ecuaciones en la forma

$$\lambda^{p(m-1)}(\xi + \omega^i\upsilon) = \delta_i(\xi + \upsilon)\gamma_i^p, \qquad i = 1, \dots, p-1, \tag{17.14}$$

donde hay que tener cuidado, porque estamos llamando $\gamma_i = \alpha_i/N(\beta_0)$ a un número ciclotómico que no es necesariamente entero. Ahora usamos la identidad

$$(\xi + \omega\upsilon)(1 + \omega) - (\xi + \omega^2\upsilon) = \omega(\xi + \upsilon),$$

que se comprueba fácilmente operando el miembro izquierdo. Multiplicamos por $\lambda^{p(m-1)}$:

$$\lambda^{p(m-1)}(\xi + \omega\upsilon)(1 + \omega) - \lambda^{p(m-1)}(\xi + \omega^2\upsilon) = \lambda^{p(m-1)}\omega(\xi + \upsilon),$$

y usamos (17.14) para $i = 1, 2$:

$$\delta_1(\xi + \upsilon)\gamma_1^p(1 + \omega) - \delta_2(\xi + \upsilon)\gamma_2^p = \lambda^{p(m-1)}\omega(\xi + \upsilon).$$

Simplificamos $\xi + \upsilon$ y multiplicamos por $\delta_1(1 + \omega)^{-1}$ (recordemos que $\omega + 1$ es una unidad):

$$\gamma_1^p - \delta_1^{-1}(1 + \omega)^{-1}\delta_2\gamma_2^p = \lambda^{p(m-1)}\omega\delta_1^{-1}(1 + \omega)^{-1}.$$

Llamamos $\zeta' = N(\mathfrak{b}_0)$, multiplicamos por ζ'^p y agrupamos las unidades:

$$\alpha_1^p + \eta\alpha_2^p = \delta'\lambda^{p(m-1)}\zeta'^p.$$

Todos los números que aparecen en esta ecuación son enteros ciclotómicos, y sabemos que α_1, α_2, ζ' no son divisibles entre λ, así como que δ' y η son unidades. Si llamamos $\xi'_1 = \alpha_1$ queda:

$$\xi'^5 + \eta\alpha_2^p = \delta'\lambda^{p(m-1)}\zeta'^p. \tag{17.15}$$

Esta ecuación es casi igual que (17.12), pero con $m - 1 > 0$ en lugar de m. Para que sea igual y tener con ello la contradicción que buscamos sólo nos falta probar que $\eta = \eta'^p$, pues entonces podremos llamar $\upsilon' = \eta'\alpha_2$, y tendremos

$$\xi'^p + \upsilon'^p = \delta'\lambda^{p(m-1)}\zeta'^p,$$

que ya es exactamente como (17.12) y nos da la contradicción que prueba que el caso II también es imposible.

En este paso final de la demostración usaremos la segunda condición del enunciado. Para ello observamos que $p(m-1) \geq p$, pues ya hemos visto que $m > 1$, por lo que el miembro derecho de (17.15) contiene al menos una potencia λ^{p-1}, luego es divisible entre p, es decir,

$$\xi'^p + \eta\alpha_2^p \equiv 0 \pmod{p}.$$

Por otro lado, por 17.25, toda potencia p-ésima es congruente con un entero racional módulo p, por lo que $u+\eta v \equiv 0 \pmod{p}$, para ciertos enteros racionales u y v. Además $p \nmid v$, pues en caso contrario $\lambda \mid p \mid \alpha_2^p$ y $\lambda \mid \alpha_2$, lo cual es falso. Por consiguiente, existe un entero w tal que $vw \equiv 1 \pmod{p}$, de donde $\eta \equiv -wu \pmod{p}$. Por hipótesis tenemos ahora que η es una potencia quinta, como teníamos que probar. ∎

Nota Probablemente al lector le habrá llamado la atención que hayamos usado letras góticas $\mathfrak{m}, \mathfrak{p}, \mathfrak{c}$, etc. para representar ciertos enteros ciclotómicos. Lo hemos hecho así para enfatizar que, en realidad, el argumento que hemos dado es válido en un contexto mucho más general.

Puede probarse que los únicos anillos de enteros ciclotómicos de orden primo con factorización única son los correspondientes a los primos

$$p = 3, \quad 5, \quad 7, \quad 11, \quad 13, \quad 19.$$

Sin embargo, Kummer demostró que todos ellos tienen una factorización única ideal análoga a la que hemos probado para cuerpos cuadráticos, de modo que puede definirse igualmente un grupo de clases de similitud de ideales que resulta ser finito, por lo que está definido el número de clases h de cada anillo de enteros ciclotómicos (de hecho, esto es cierto para cualquier anillo de enteros algebraicos), de modo que la factorización única (real) equivale a que $h = 1$.

Pero resulta que la prueba que hemos dado no requiere realmente $h = 1$, sino meramente que $p \nmid h$, condición que se cumple, no para todos, pero sí para

muchos más primos que los seis que tienen factorización única. Por ejemplo, de los 45 primos impares menores que 200 la cumplen todos excepto ocho:

$$37,\quad 59,\quad 67,\quad 101,\quad 103,\quad 131,\quad 149,\quad 157.$$

En la prueba bajo la hipótesis $p \nmid h$, los enteros ciclotómicos que hemos nombrado con letras góticas pasan a ser ideales de $\mathbb{Z}[\omega]$.

En cuando a la segunda hipótesis, que nosotros hemos demostrado explícitamente en 17.26 para $p = 5$, Kummer demostró —aunque no es trivial— que se cumple siempre que $p \nmid h$, por lo que ésta es en realidad la única hipótesis necesaria para probar el Último Teorema de Fermat para un exponente p (mediante el argumento de Kummer). ∎

17.8 Enteros ciclotómicos y sumas de Gauss

En esta sección demostraremos la fórmula (17.1) que hemos discutido en la introducción, que, para cada primo $p \equiv 1 \pmod 4$, relaciona el número de clases h del anillo de enteros del cuerpo $K = \mathbb{Q}\big(\sqrt{p}\,\big)$ con su unidad fundamental ϵ, a la vez que determina el resto módulo p del factorial de $(p-1)/2$.

El ingrediente principal de la prueba es la fórmula para el número de clases de K que hemos obtenido analíticamente en [ITAn]. En lugar de usar la expresión recogida en el teorema [ITAn 11.7], en términos de senos, consideraremos las dos determinaciones de $L(1, \chi_K)$ obtenidas en [ITAn 8.34] y en [ITAn (11.3)]. La primera (evaluando la suma de Gauss) es

$$L(1, \chi_K) = -\frac{1}{\sqrt{p}} \sum_{k=1}^{p-1} \left(\frac{k}{p}\right) \log(1 - \omega^{-k}),$$

mientras que la segunda es

$$L(1, \chi_K) = \frac{2h \log \epsilon}{\sqrt{p}}.$$

Al combinarlas obtenemos que

$$2h \log \epsilon = -\sum_{k=1}^{p-1} \left(\frac{k}{p}\right) \log(1 - \omega^{-k}).$$

En el sumatorio podemos cambiar k por $-k$, y así, como $(-1/p) = 1$,

$$\log \epsilon^{2h} = -\sum_{k=1}^{p-1} \left(\frac{k}{p}\right) \log(1 - \omega^{k}).$$

Aplicando la función exponencial:

$$\epsilon^{2h} = \prod_{k=1}^{p-1} (1 - \omega^k)^{-(k/p)}.$$

Para eliminar el cuadrado multiplicamos por $p = \prod_{k=1}^{p-1}(1-\omega^k)$, con lo que

$$p\epsilon^{2h} = \prod_{(k/p)=-1}(1-\omega^k)^2.$$

Por lo tanto:

$$\sqrt{p}\,\epsilon^h = \prod_{(k/p)=-1}(1-\omega^k), \tag{17.16}$$

donde hemos tenido en cuenta que ambos miembros son positivos, pues, como $(k/p) = (-k/p)$, en el producto podemos agrupar pares de factores conjugados

$$(1-\omega^k)(1-\omega^{-k}) = |1-\omega^k|^2 > 0.$$

En este punto vamos a aprovechar que estamos trabajando con elementos del anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos de orden p. En primer lugar observamos que, el hecho de que la suma de Gauss $G(p) = \sqrt{p}$ (en realidad, nos basta con el hecho más fácil de probar $G(p) = \pm\sqrt{p}$), podemos concluir que $\sqrt{p} \in \mathbb{Z}[\omega]$, luego $\mathbb{Q}\big(\sqrt{p}\,\big) \subset \mathbb{Q}(\omega)$.

Fijemos ahora un resto no cuadrático n_0 módulo p y consideremos la conjugación σ_{n_0}. Entonces, usando el teorema 9.22:

$$\sigma_{n_0}(\sqrt{p}) = \sigma_{n_0}(G(p)) = \sum_{k=1}^{p-1}\left(\frac{k}{p}\right)\zeta^{n_0 k} = G_{n_0}(p) = \left(\frac{n_0}{p}\right)G(p) = -G(p).$$

Esto significa que la conjugación σ_{n_0}, cuando actúa sobre los elementos de $\mathbb{Q}\big(\sqrt{p}\,\big)$, coincide con la conjugación (no trivial) de este cuerpo cuadrático, determinada por $\sqrt{p} \mapsto -\sqrt{p}$. Sabiendo esto, aplicamos σ_{n_0} a (17.16):

$$-\sqrt{p}\;\bar{\epsilon}^h = \prod_{(k/p)=-1}(1-\omega^{n_0 k}),$$

donde $\bar{\epsilon}$ representa el conjugado de ϵ en $\mathbb{Q}\big(\sqrt{p}\,\big)$. Ahora bien, cuando k recorre los restos no cuadráticos módulo p, tenemos que $n_0 k$ recorre los restos cuadráticos, luego la última igualdad es equivalente a

$$-\sqrt{p}\;\bar{\epsilon}^h = \prod_{(k/p)=1}(1-\omega^k).$$

Multiplicando esta igualdad con (17.16) queda:

$$-p\,\mathrm{N}(\epsilon)^h = \prod_{k=1}^{p-1}(1-\omega^k) = p,$$

luego $\mathrm{N}(\epsilon)^h = -1$.

Recogemos en un teorema lo que hemos obtenido:

Teorema 17.28 *Si $p \equiv 1$ (mód 4) es un número primo, la unidad fundamental ϵ del anillo de enteros algebraicos del cuerpo $\mathbb{Q}(\sqrt{p}\,)$ cumple $\mathrm{N}(\epsilon) = -1$ y el número de clases h es impar.*

Para probar la fórmula (17.1) necesitamos otra aplicación crucial de la aritmética del anillo $\mathbb{Z}[\omega]$. En realidad se trata de un hecho que hemos demostrado ya sin mencionar explícitamente los enteros ciclotómicos.

En efecto, en la demostración del teorema 7.10, en el que hemos determinado el signo de las sumas de Gauss cuadráticas de orden primo, hemos estado manejando "entre líneas" el anillo $\mathbb{Z}[\omega]$ y su aritmética. Técnicamente hemos trabajado en el anillo de polinomios $\mathbb{Z}_p[x]$, donde hemos considerado congruencias módulo potencias del polinomio $\pi = x - \bar{1}$, pero sucede que eso es equivalente a trabajar en el anillo $\mathbb{Z}[\omega]$ y tomar congruencias módulo potencias del primo $\lambda = \omega - 1$.

En efecto, consideremos concretamente la congruencia (7.7), que es la que nos interesa:

$$\bar{G}(\pi + 1) \equiv -\frac{\pi^{(p-1)/2}}{((p-1)/2)!} \text{ (mód } \pi^{(p+1)/2}).$$

Recordemos que $\pi + 1 = x$ y que

$$G(x) = \sum_{k=1}^{p-1} \left(\frac{k}{p}\right) x^k \in \mathbb{Z}[x],$$

de modo que $\bar{G}$ es el polinomio en $\mathbb{Z}_p[x]$ que resulta de sustituir los coeficientes de G por sus restos módulo p. Explícitamente, la congruencia significa que existe un polinomio $\bar{u}(x) \in \mathbb{Z}_p[x]$ tal que

$$\tfrac{p-1}{2}!\, \bar{G}(x) = -(x - \bar{1})^{(p-1)/2} + \bar{u}(x)(x-1)^{(p+1)/2}.$$

Más explícitamente aún, en términos de polinomios de $\mathbb{Z}[x]$, existe otro polinomio $v(x) \in \mathbb{Z}[x]$ tal que

$$\tfrac{p-1}{2}!\, G(x) = -(x-1)^{(p-1)/2} + u(x)(x-1)^{(p+1)/2} + pv(x).$$

Si ahora evaluamos estos polinomios en $\omega = e^{2\pi i/p}$, obtenemos que

$$\tfrac{p-1}{2}!\, G(p) = -\lambda^{(p-1)/2} + u(\omega)\lambda^{(p+1)/2} + pv(\omega),$$

pero sabemos que $\lambda^{p-1} \mid p$, por lo que

$$\tfrac{p-1}{2}!\, G(p) \equiv -\lambda^{(p-1)/2} \text{ (mód } \lambda^{(p+1)/2}).$$

Esta congruencia es esencialmente la misma que (7.7). En el caso que nos ocupa, en que $p \equiv 1$ (mód 4), podemos evaluar la suma de Gauss y concluir que

$$\sqrt{p} \equiv -\frac{1}{\frac{p-1}{2}!}\lambda^{(p-1)/2} \text{ (mód } \lambda^{(p+1)/2}).$$

Ahora introducimos esta congruencia en (17.16), con lo que

$$\prod_{(k/p)=-1}(1-\omega^k) \equiv -\tfrac{\epsilon^h}{\frac{p-1}{2}!}\lambda^{(p-1)/2} \text{ (mód } \lambda^{(p+1)/2}).$$

Equivalentemente:

$$\epsilon^h\lambda^{(p-1)/2} \equiv -\tfrac{p-1}{2}! \prod_{(k/p)=-1}(1-\omega^k) \text{ (mód } \lambda^{(p+1)/2}).$$

Ya hemos visto que

$$1-\omega^k = (1-\omega)(\omega^{k-1}+\cdots+\omega+1),$$

luego $\lambda \mid (1-\omega^k)$, así que podemos escribir

$$\epsilon^h \equiv -\tfrac{p-1}{2}! \prod_{(k/p)=-1}\frac{1-\omega^k}{\lambda} \text{ (mód } \lambda).$$

Además, truncando el desarrollo del binomio obtenemos que:

$$1-\omega^k = 1-(1+\lambda)^k \equiv 1-(1+k\lambda) \equiv -k\lambda \text{ (mód } \pi)^2,$$

luego

$$\epsilon^h \equiv -\tfrac{p-1}{2}! \prod_{(k/p)=-1}(-k) = -\tfrac{p-1}{2}! \prod_{(k/p)=-1} k \text{ (mód } \lambda),$$

donde hemos usado que el número de factores es $(p-1)/2$, que es par, porque $p \equiv 1$ (mód 4).

Ahora bien, el producto de los restos cuadráticos módulo p es

$$1^2\cdot 2^2\cdots\Big(\frac{p-1}{2}\Big)^2 = \Big(\frac{p-1}{2}!\Big)^2 \equiv -1 \text{ (mód } p)$$

(véase la introducción al capítulo XI de [ITAn]) y como, por el teorema de Wilson,

$$\prod_{(k/p)=-1} k \prod_{(k/p)=1} k \equiv (p-1)! \equiv -1 \text{ (mód } p),$$

concluimos que

$$\prod_{(k/p)=-1} k \equiv 1 \text{ (mód } p).$$

Por lo tanto,

$$\epsilon^h \equiv -\tfrac{p-1}{2}! \text{ (mód } \lambda).$$

Para concluir observamos que si $\epsilon = (u_0+v_0\sqrt{p})/2$, como $\lambda \mid \sqrt{p}$ (porque $\lambda \mid p = (\sqrt{p})^2$ y es primo), vemos que $2\epsilon \equiv u_0$ (mód λ), e igualmente tenemos que $2\bar{\epsilon} \equiv u_0$ (mód λ), luego $2\bar{\epsilon} \equiv 2\epsilon$ (mód λ) y también $\bar{\epsilon} \equiv \epsilon$ (mód λ), luego, $-1 = \mathrm{N}(\epsilon) = \epsilon\bar{\epsilon} \equiv \epsilon^2$ (mód λ). Así, como h es impar, llegamos a que

$$2\tfrac{p-1}{2}! \equiv -2\epsilon^h = -(\epsilon^2)^{(h-1)/2}2\epsilon \equiv (-1)^{(h+1)/2}u_0 \text{ (mód } \lambda),$$

pero esto significa que λ divide a la diferencia de los dos miembros, que es un entero racional, luego, tomando norma, concluimos que p también la divide, es decir, que

$$2^{\frac{p-1}{2}}! \equiv (-1)^{(h+1)/2} u_0 \text{ (mód } p),$$

Si expresamos $\epsilon = u + v\sqrt{p}$, de modo que $u_0 = 2u$, con u entero o semientero, llegamos a la fórmula (17.1), que queríamos probar. ∎

Índice de Materias