

Carlos Ivorra Castillo

**INTRODUCCIÓN AL
CÁLCULO DIFERENCIAL**

... quien, con su fuerza mental casi divina, siguiendo la luz de su matemática, demostró el primero los movimientos y figuras de los planetas, los senderos de los cometas y el flujo y reflujos del Océano. . .

EPITAFIO DE NEWTON

Índice General

Introducción	vii
Capítulo I: Derivadas	1
1.1 La derivada de una función	3
1.2 Cálculo de derivadas	11
1.3 Máximos y mínimos	22
1.4 Propiedades de las funciones derivables	26
1.5 La regla de L'Hôpital	43
1.6 Continuidad de la derivada	50
1.7 Ecuaciones diferenciales I	52
1.8 La diferencial de una función	63
Capítulo II: El cálculo integral	67
2.1 Integrales definidas e indefinidas	69
2.2 Cálculo de primitivas	79
2.3 Las leyes de Kepler	92
2.4 El teorema de aproximación de Weierstrass	95
2.5 La longitud de un arco	98
2.6 Cálculo de volúmenes	108
2.7 Áreas de sólidos de revolución	117
2.8 Ecuaciones diferenciales II	124
2.9 El teorema de Arzelà	139
Capítulo III: Integrales impropias	143
3.1 Definición y ejemplos	144
3.2 Criterios de convergencia	151
3.3 Derivación de integrales paramétricas	160
3.4 Integrales dobles	166
3.5 La transformada de Laplace	171
3.6 La función factorial	188
Capítulo IV: La dinámica newtoniana	201
4.1 Cinemática	201
4.2 Las leyes de Newton	221
4.3 La ley de gravitación universal	236

4.4	Sistemas de referencia no inerciales	251
Capítulo V: Diferenciabilidad		281
5.1	Funciones diferenciables	282
5.2	Ejemplos y resultados fundamentales	291
5.3	Teoría básica	316
5.4	La energía mecánica	325
5.5	El problema de la braquistócrona	344
Capítulo VI: Integrales y funciones elípticas		353
6.1	Las integrales de primera y segunda especie	353
6.2	La lemniscata de Bernoulli	364
6.3	Las funciones elípticas lemniscáticas	369
6.4	Las funciones elípticas de Jacobi	376
6.5	El péndulo simple	388
6.6	El péndulo esférico	393
6.7	La media aritmético-geométrica	402
Capítulo VII: Introducción al análisis complejo		415
7.1	Funciones holomorfas	416
7.2	La integral curvilínea	421
7.3	Propiedades de las funciones holomorfas	433
7.4	El teorema de Cauchy	448
Apéndice A: Geometría analítica tridimensional		453
Apéndice B: Ejemplos de curvas		467
B.1	La cicloide	467
B.2	La cardioide	472
B.3	La nefroide	474
B.4	La tractriz	477
B.5	La lemniscata de Bernoulli	481
B.6	Los óvalos de Cassini	483
Apéndice C: Cálculo de primitivas		489
C.1	Integrales inmediatas	489
C.2	Funciones racionales	492
C.3	Integración por partes	498
C.4	Integración por sustitución	503
Índice de Materias		509

Introducción

Aristóteles En 336 a.C. regresó a Atenas Aristóteles, un antiguo alumno de la Academia de Platón. Había estado fuera unos años, ejerciendo de preceptor de Alejandro Magno y, en lugar de reincorporarse a la Academia, fundó su propia escuela, que fue conocida como el Liceo. Aristóteles se interesó por prácticamente todas las ramas del saber, con menos interés que su maestro por la geometría y más por las ciencias naturales. En 323, cuando tendría unos 60 años, dejó el Liceo en manos de su discípulo Teofrasto, bajo cuya dirección llegó a tener más de 2 000 alumnos, y su biblioteca llegó a contar con más de 10 000 papiros, que incluían más de 157 títulos de Aristóteles y otros 225 de Teofrasto. Actualmente se conservan poco más de una treintena de libros atribuidos a Aristóteles, pero están redactados con tal aire académico, con tanta pompa y tanta seguridad, que cuando fueron redescubiertos en el Renacimiento fueron tomados por la perfección del saber, y pronto se extendió por Occidente la idea de que contradecir al Estagirita era más o menos como contradecir la Biblia.

Entre los científicos griegos que le siguieron no fue tan venerado como lo sería posteriormente, y no dudaron en contradecirle cuando lo estimaron oportuno. Por ejemplo, Aristóteles había afirmado que el Sol y los planetas giran alrededor a la Tierra en órbitas circulares a velocidad constante: no podía ser de otra manera, pues el Cielo es perfecto y la circunferencia es la curva perfecta, y tampoco estaría bien visto en un planeta ir cambiando de velocidad según el momento. Pero esto era incoherente con las observaciones: ya desde mucho antes los astrónomos babilónicos habían registrado las posiciones de los planetas en la esfera celeste en días sucesivos durante largos periodos de tiempo y habían constatado que de vez en cuando pasan de avanzar a retroceder y volver a avanzar dibujando una elipse sobre las estrellas. Esto no era posible si se movían como Aristóteles afirmaba. Poco después de la muerte de Aristóteles, antes del 300 a.C., Apolonio de Perge tuvo la idea de explicar este movimiento suponiendo que los planetas describen órbitas circulares alrededor de puntos que a su vez describen órbitas circulares alrededor de la Tierra (algo así como la Luna que gira alrededor de la Tierra que gira alrededor del Sol, pero sin Tierra). Estas órbitas sobre órbitas se conocieron como *epiciclos*. Ya en el siglo II a.C., Hiparco desarrolló esta idea, pero en el siglo II d.C. Ptolomeo se dio cuenta de que su modelo no encajaba con las observaciones babilónicas más antiguas, así que, en un auténtico *tour de force*, empleó toda serie de trucos similares para construir un modelo planetario que se ajustase al milímetro a las observaciones.

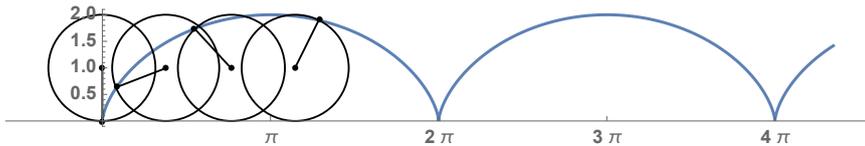
Para ello tuvo que suponer epiciclos sobre epiciclos sobre epiciclos, así como que la velocidad de giro no era constante medida desde la Tierra, sino desde otro punto asociado al planeta al que llamó “ecuante”. Con estos ajustes *ad hoc*, logró un sistema astronómico que sería aceptado durante siglos por su eficiencia en la predicción de las posiciones de los astros, aunque fuera artificioso y matemáticamente complejo. Este sistema quedó plasmado en su *Almagesto*, terminado sobre 150 d.C.

Copérnico Ya en 270 a.C., Aristarco de Samos había propuesto que la Tierra y los planetas giran alrededor del Sol, si bien la idea nunca acabó de cuajar entre los griegos, y cuando en 1531 el astrónomo polaco Nicolás Copérnico trató de crear un sistema heliocéntrico alternativo al sistema ptolemaico, se encontró con que no podía prescindir de los epiciclos, aunque sí de los ecuantos, y, si bien en su conjunto el modelo resultó conceptualmente más simple, no dejaba de ser artificioso. Copérnico se mostró reacio a publicar su teoría temiendo que no sería aceptada, pero varios astrónomos lo animaron a hacerlo, y así en 1542 dio a la imprenta su famoso *De revolutionibus orbium caelestium*, aunque el autor moriría poco después, ese mismo año.

La obra era bastante técnica, así que pasó inadvertida salvo en los sectores más especializados, pero poco a poco fueron apareciendo divulgadores de la teoría heliocéntrica. No obstante, lo cierto era que, aunque existían indicios que hacían plausible el heliocentrismo, había muchos más argumentos en su contra, y no sólo “argumentos” teológicos y la veneración a la palabra de Aristóteles (que no había sostenido el sistema ptolemaico estrictamente hablando, pero interpretando sus palabras en sentido amplio, podían considerarse consistentes con él), sino que también “la física elemental” parecía contradecir el heliocentrismo. Para que finalmente fuera aceptado, tuvo que venir alguien a revolucionar “la física elemental”.

Galileo I En 1581, un joven de 17 años observó cómo un sacristán encendía la lámpara que colgaba de una cadena de la cúpula de la catedral de Pisa. Para ello tuvo que acercarla y, cuando la soltó, la lámpara osciló durante un tiempo, y el muchacho tuvo la sensación de que el ritmo con que se balanceaba era siempre el mismo, aunque la amplitud del balanceo fuera disminuyendo. Para comprobarlo, usó su propio pulso para medir el tiempo de las oscilaciones, y el procedimiento, aunque muy inexacto, pareció confirmar su impresión. Deseoso de comprobarlo con más precisión, se fue a su casa, ató dos esferas a dos cuerdas de la misma longitud, las puso a oscilar con amplitudes diferentes y pidió a un amigo que contara las oscilaciones de una mientras él contaba las de la otra, y, en efecto, comprobó que las dos marcaban el mismo ritmo. El joven Galileo Galilei había descubierto lo que llamó la *isocronía* de los péndulos, y esto fue el punto de partida de muchas otras reflexiones sobre la naturaleza del movimiento, así como de la importancia de la experimentación para comprenderla correctamente. Así, por ejemplo, experimentando con cuerdas de distintas longitudes, llegó a la conclusión de que el periodo de oscilación de un péndulo, es decir, el tiempo que tarda en volver a su posición inicial, es proporcional a la raíz cuadrada de la longitud de la cuerda.

En 1589 el duque Fernando I de Medici lo nombró catedrático de matemáticas de la Universidad de Pisa. Uno de sus objetos de estudio en este periodo que tiene relación con el contenido de este libro fue el descubrimiento de una curva que aparecía ya en un libro de principios de siglo del matemático francés Charles de Bovelles, pero a la que Galileo dio el nombre con que se la conoce actualmente: la *cicloide*. Puede definirse como la trayectoria que sigue un clavo clavado en el borde de una rueda cuando ésta gira:



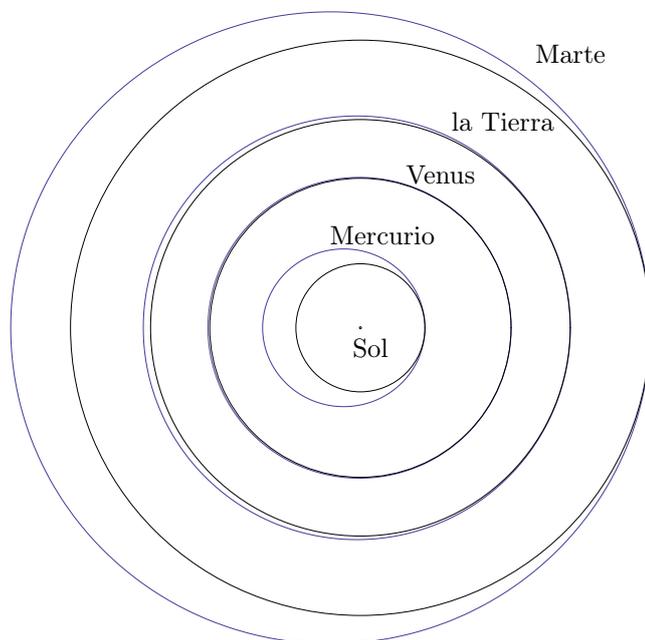
Galileo la usó para diseñar puentes. En 1592 se trasladó a la Universidad de Padua, como profesor de geometría, mecánica y astronomía. Allí, en 1599 trató en vano de calcular el área limitada por un arco de cicloide. Llegó incluso a dibujar la rueda y la cicloide sobre una lámina de metal, recortarlas y pesar las piezas para analizar su proporción, y vio que ésta era aproximadamente 3 : 1, pero no llegó a ninguna conclusión definitiva.

Kepler Mientras Galileo enseñaba a sus alumnos el sistema ptolemaico, el heliocentrismo recibió un nuevo impulso. En 1601 murió el astrónomo danés Tycho Brahe, y fue sucedido como matemático imperial del emperador Rodolfo II de Austria por un astrólogo y astrónomo alemán llamado Johannes Kepler, que estaba convencido de que Dios no podía haber hecho el cosmos tan complicado como lo pintaba Ptolomeo, y que una teoría heliocéntrica adecuada demostraría la simplicidad y la elegancia del plan divino. Usando datos muy precisos sobre la posición de Marte que había recopilado Brahe a lo largo de su vida, Kepler descartó unos 40 intentos fallidos de describir su órbita con absoluta precisión y simplicidad, hasta que finalmente, en 1604, como último recurso desesperado, probó a ajustar los datos a una elipse y vio que cuadraban perfectamente. Así llegó a un principio fundamental:

Primera ley de Kepler *Los planetas siguen órbitas elípticas con el Sol en uno de sus focos.*

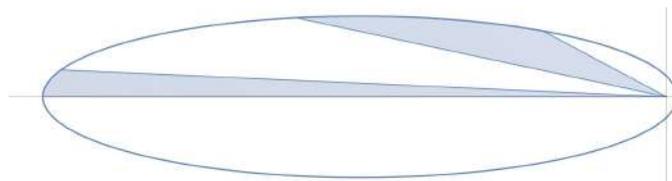
Con esta hipótesis, el movimiento planetario se explicaba sin necesidad de epiciclos, ecuantes ni artificio alguno. En realidad las órbitas planetarias son elipses con excentricidad muy pequeña, es decir, muy parecidas a circunferencias. La figura siguiente muestra a escala las órbitas de los planetas más cercanos al Sol junto con la órbita circular de radio igual al perihelio. En la figura los perihelios están alineados, pero esto no sucede realmente. Vemos que Mercurio y Marte tienen una excentricidad mucho mayor que la Tierra, y la de ésta es mayor a su vez que la de Venus, que es casi inapreciable.

En cuanto a las velocidades de rotación, Kepler comprobó que se cumple una ley que ya había estado considerando en sus intentos previos.



Segunda ley de Kepler *El radio que une un planeta con el Sol barre áreas iguales en tiempos iguales.*

Para visualizar el contenido geométrico de esta ley conviene considerar una órbita de excentricidad apreciable, como la del cometa Halley (pues sucede que todos los cuerpos que orbitan de forma estable alrededor del Sol —no sólo los planetas— siguen órbitas elípticas). La figura muestra a escala la órbita del cometa, y lo que afirma la segunda ley de Kepler es que éste recorre en el mismo tiempo los dos arcos señalados, pues las áreas sombreadas son iguales.



Esto implica que un planeta (o cometa, etc.) se mueve más rápido cuanto más cerca está del Sol que cuando está lejos, pues cuando está lejos, avanzando menos, barre un área mayor que cuando está cerca. Kepler incluyó sus dos leyes en su tratado *Astronomia noua*, publicado en 1609. En 1618 Kepler descubrió una tercera ley sobre el movimiento planetario:

Tercera ley de Kepler *Los cuadrados de los períodos de revolución de los planetas son proporcionales a los cubos de los semiejes mayores de sus órbitas.*

Una consecuencia notable de esta ley es que proporciona la escala del sistema solar, pues los periodos de revolución de los planetas son fáciles de medir. Por ejemplo, el periodo de Venus es $T_v = 224.7$ días y el de la Tierra es $T_t = 365.25$ días y, como

$$\frac{T_v^2}{a_v^3} = \frac{T_t^2}{a_t^3},$$

tenemos que

$$\frac{a_v}{a_t} = \sqrt[3]{\frac{T_v^2}{T_t^2}} = 0.7233.$$

La distancia media de la Tierra al Sol es lo que se conoce como una *unidad astronómica*. En estos términos, hemos calculado que Venus está a 0.7233 unidades astronómicas del Sol.

En su *De revolutionibus*, de 1543, Copérnico ya había calculado las distancias al Sol de los planetas conocidos midiendo el ángulo entre cada uno de ellos y el Sol en el momento en que formaban con la Tierra un triángulo rectángulo con ángulo recto en el Sol. Su estimación para el caso de Venus fue que está a 0.719 unidades astronómicas del Sol.

Así pues, los astrónomos sólo necesitaban determinar la distancia de un planeta al Sol para conocer de hecho la distancia al Sol de todos los planetas. La unidad astronómica ha sido determinada con gran precisión mediante mediciones por radar desde sondas espaciales de la distancia de la Tierra a Venus o a Marte:

$$1 \text{ ua} = 149\,597\,870\,700 \text{ m}.$$

Las estimaciones de los astrónomos antiguos eran bastante pobres. Por ejemplo, Ptolomeo la estimó en un 5% de su valor real, y su valor era mayor que los de sus predecesores. La primera medición fiable de la unidad astronómica la llevaron a cabo Cassini y Richer en 1672 (véase la sección 3.5 de [IGE]).

Por otra parte, conviene recordar que la unidad astronómica es la distancia media de la Tierra al Sol, pero el semieje mayor de la Tierra mide 1.00000261 unidades astronómicas, luego a menudo no es necesario hacer la distinción.

Otro ejemplo: el cometa Halley tarda 75.3 años en dar una vuelta alrededor del Sol, por lo que, según la tercera ley de Kepler, el semieje mayor de la órbita del cometa Halley cumple

$$\frac{a_h}{a_t} = \sqrt[3]{\frac{T_h^2}{T_t^2}} = \sqrt[3]{\frac{75.3^2}{1^2}} \approx 17.83 \text{ ua}.$$

El valor observado es de 17.8576 ua.

Con las leyes de Kepler, la teoría heliocéntrica había dejado de ser una teoría formalmente análoga a la geocéntrica para convertirse en una alternativa mucho más simple y natural, aunque Kepler no podía dar ninguna razón —aparte de la concordancia con las observaciones— de por qué los planetas tenían que seguir órbitas elípticas o por qué se dedicaban a barrer las áreas de sus elipses.

Galileo II En 1609 Galileo tuvo noticia de un invento neerlandés llamado “telescopio” que permitía ver estrellas invisibles a simple vista. A partir de unas pocas indicaciones sobre el principio de su funcionamiento, se dedicó a construir sus propios telescopios cada vez más precisos, y en 1610 ya había construido unos 60. Todos ellos habían causado sensación en la corte, pero sólo unos pocos servían para hacer observaciones astronómicas. Ese año publicó su *Sidereus nuntius*, donde informaba de algunos de sus descubrimientos, como que la Luna no es una esfera lisa, perfecta, sino que tiene valles y montañas como la Tierra, o que Júpiter tiene al menos cuatro lunas que orbitan a su alrededor. Posteriormente, ese año, descubrió también que Venus presenta fases, como la Luna, cosa que difícilmente puede explicarse desde el geocentrismo, y que el Sol tiene manchas que a su vez indican que gira sobre sí mismo.

Estos descubrimientos y las demostraciones que organizaba le valieron una gran admiración y popularidad, y ese año se trasladó a Florencia. Sin embargo, también se ganó muchos enemigos, principalmente entre los académicos que veían cuestionada su reputación y sus conocimientos por las novedades que sostenía Galileo y su defensa del heliocentrismo. Entre sus réplicas no tardaron en aparecer argumentos teológicos y la Iglesia intervino. En 1616 la Santa Inquisición lo hizo comparecer para examinar sus teorías astronómicas y concluyó que la teoría copernicana era “insensata, un absurdo en filosofía y una herejía”. Esto no tuvo consecuencias graves: Galileo no fue condenado, pero “se le rogó” que presentara sus teorías como meras hipótesis equivalentes al geocentrismo.

Pero los argumentos que se daban contra el heliocentrismo no eran meramente teológicos o filosóficos, sino que también había argumentos físicos que no era obvio cómo rebatir, y Galileo tendría que meditar mucho tiempo sobre ellos para refutarlos debidamente.

Por ejemplo, uno de ellos consistía en que si la Tierra gira alrededor de sí misma a razón de una vuelta diaria —como supone la teoría heliocéntrica para explicar el movimiento diario del Sol—, cualquier punto de la superficie terrestre se estaría moviendo de oeste a este a una velocidad enorme. Más concretamente, considerando que el radio de la Tierra es de unos 6 300 km, un punto situado en Florencia tendría que estar moviéndose a unos 330 m/s, por lo que si lanzamos una piedra desde lo alto de una torre y tarda 3 s en llegar al suelo, no tendría que caer en la base de la torre, sino unos $3 \cdot 330 = 990$ m hacia el oeste. Igualmente se argumentaba que una flecha disparada hacia el este debería impactar en la cara del arquero, etc.

No fue hasta 1632 cuando Galileo publicó sus refutaciones a estos y otros argumentos en su *Diálogo sobre los dos máximos sistemas del mundo, ptolemaico y copernicano*. El mayor hito conceptual que contiene es lo que hoy se conoce como *principio de relatividad de Galileo*, el cual a su vez depende de otro hecho fundamental que en términos modernos puede enunciarse así:

Todo movimiento es relativo a un sistema de referencia.

Por ejemplo, un pasajero sentado en un asiento de un tren en movimiento puede distinguir entre los pasajeros que están en reposo, sentados en otros asientos, y los que se están moviendo por el pasillo del tren. En este caso, estamos

hablando de movimiento *relativo al tren*. Sin embargo, todos los pasajeros, tanto los que están en reposo como los que están en movimiento respecto del tren, están en movimiento respecto, por ejemplo, de un pasajero que los observa por las ventanillas desde un andén de una estación. El tren y la estación son dos *sistemas de referencia* respecto a los que se puede describir el movimiento de un cuerpo, de forma que un mismo cuerpo puede estar en reposo respecto de uno y en movimiento respecto del otro.

Quizá el lector piense que esto es jugar con las palabras, en el sentido de que los pasajeros que están en reposo respecto del tren, “en realidad” están en movimiento, porque el tren se mueve, pero la esencia del principio que acabamos de formular es que a “el tren se mueve” hay que añadir inevitablemente un “respecto de la Tierra”, porque no tiene sentido decir “se mueve” sin especificar respecto a qué. Esto es una consideración previa al principio de relatividad de Galileo propiamente dicho, que afirma lo siguiente:

Principio de relatividad de Galileo *Si un sistema de referencia se mueve respecto de otro en línea recta con velocidad constante, entonces las leyes del movimiento son exactamente las mismas cuando el movimiento se describe con respecto a uno o a otro.*

Por ejemplo, según este principio, si alguien está en un tren sin ventanillas, no puede hacer ningún experimento mecánico (dejar caer una bola y observar su movimiento, etc.) que le permita distinguir si el tren está parado ante una estación o se está moviendo por una vía recta a velocidad constante. (Galileo no sabía lo que era un tren, pero formuló esto mismo hablando de un pasajero metido en la bodega de un barco).

Notemos que esto no es un principio metafísico, sino un principio físico empíricamente constatable: puede observarse que si, por ejemplo, dejamos caer una bola a bordo de un tren que se mueve a 200 km/h en línea recta, veremos caer la bola verticalmente, y no en dirección a la cola del tren, y si la lanzamos hacia la cabeza del tren, la veremos moverse a la misma velocidad con que la veríamos moverse si la hubiéramos lanzado estando parados en la estación. Este principio empíricamente constatable explica por qué la Tierra puede girar alrededor de sí misma sin que las flechas que disparan los arqueros se les estampen en la cara, etc.

Esto no es aplicable al caso de que el tren no se mueva a velocidad constante. Todos sabemos lo que sucede cuando un tren, o un coche, acelera o frena, que nos vemos empujados hacia atrás o hacia adelante según el caso y así podemos constatar que el vehículo está en movimiento. Tampoco se aplica si en lugar de estar dentro del tren estamos sobre su techo, porque entonces el aire que está en reposo para los observadores situados en tierra es para nosotros un vendaval a 200 km/h que nos empujará hacia la cola del tren.

El diálogo de Galileo motivó la intervención de la Santa Inquisición. En el diálogo Galileo aportaba nuevas pruebas que refutaban el geocentrismo, contraviniendo la “recomendación” de presentar la teoría heliocéntrica como una mera hipótesis no demostrable. Pero más determinante fue que sus enemigos

difundieron la idea de que el personaje geocentrista del diálogo era una caricatura del papa Urbano VIII. En 1633 la Santa Inquisición abrió un proceso contra él. Bajo amenazas de ser torturado y promesas de un trato benevolente si colaboraba, Galileo confesó su culpa, abjuró del herético heliocentrismo y fue condenado a cadena perpetua, pena que el papa le conmutó inmediatamente por la de arresto domiciliario.

Para entonces Descartes tenía prácticamente terminado su *Tratado del mundo y de la luz*, en el que exponía sus teorías filosóficas y físicas, en particular un sistema heliocéntrico, pero en vista de la situación decidió no publicarlo y sólo en 1664 se decidió a darlo a la imprenta.

En el momento de su condena, Galileo tenía 69 años, y aprovechó su retiro forzoso para seguir investigando sobre el movimiento de los cuerpos, algo que podía hacer tranquilamente en sus habitaciones. Sus conclusiones las plasmó en su último libro, *Diálogos y demostraciones matemáticas sobre dos nuevas ciencias*, que completó poco antes de quedarse ciego y cuyo manuscrito llegó a los Países Bajos y a Francia, donde fue publicado en 1638. En el expone ideas nuevas que atacaban las bases de la física aristotélica. En efecto, Aristóteles había afirmado que, en el mundo sublunar, los cuerpos tienen una tendencia natural al reposo, de modo que para que un cuerpo permanezca en movimiento es necesario estar empujándolo constantemente, o al menos periódicamente (como cuando alguien viaja en patinete, impulsándose con el pie de tanto en tanto). Galileo comprendió que esto no es exactamente así, que es cierto que los cuerpos terminan parándose, pero ello se debe a que los frena el rozamiento con el suelo o con el aire, de modo que, cuanto menor sea éste, más tarda el cuerpo en pararse, por lo que si un cuerpo se moviera en el vacío, sin rozamiento alguno, su velocidad permanecería constante. Esto es lo que se conoce como el principio de inercia de Galileo:

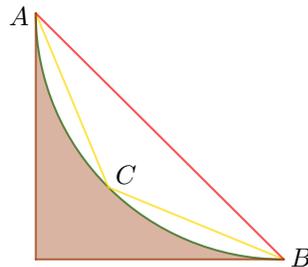
Principio de inercia de Galileo *Si un cuerpo está libre de toda influencia externa, o permanece en reposo, o se mueve en línea recta con velocidad constante.*

A decir verdad, Galileo conservó una parte del aristotelismo, porque pensó que el principio de inercia sólo era aplicable en esta forma en la Tierra, mientras que la versión para cuerpos celestes afirmaba que, igual que los cuerpos terrestres tienden a mantener su velocidad constante en línea recta, los cuerpos celestes, que están libres de la influencia gravitatoria, tienden a mantener su velocidad constante en órbitas circulares.

Pero, ciñéndonos al mundo “sublunar”, ya de joven, Galileo había constatado un principio que contradecía la física aristotélica, y es que si se deja caer dos cuerpos desde la misma altura, ambos tardan el mismo tiempo en llegar al suelo, independientemente de su peso, siempre y cuando podamos despreciar el rozamiento con el aire, que hace que una hoja de papel tarde más en caer que una piedra, debido a que puede “flotar” en el aire. En ausencia de aire, tardaría lo mismo en caer que la piedra.¹ Más precisamente, Galileo comprobó

¹En 1971 el astronauta David Scot dejó caer a la vez ante una cámara un martillo y una pluma en la Luna, y ambos llegaron al suelo al mismo tiempo.

empíricamente que la distancia que recorre un cuerpo que cae —en la medida en que se pueda despreciar el rozamiento— es proporcional al cuadrado del tiempo empleado, donde la constante de proporcionalidad es la misma para todos los cuerpos, independientemente de su peso, lo cual equivale a que su velocidad en cada instante es proporcional al tiempo transcurrido. Además comprobó que esto es así incluso si la caída se efectúa por un plano inclinado, aunque entonces cambia la constante de proporcionalidad. Sus argumentos eran algebraicos y geométricos, y le permitieron plantear y resolver algunos problemas interesantes. Por ejemplo, consideremos la figura siguiente:



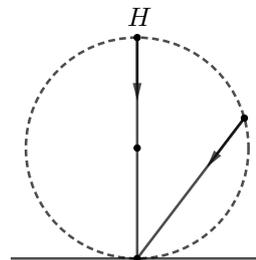
Imaginemos que la recta diagonal es una rampa muy lisa que se eleva 100 m y que dejamos caer por ella un cuerpo desde el punto A . Galileo calculó (no con estos datos precisamente) que el tiempo que tardará en llegar hasta B es de 6.39 s, mientras que si la dejamos caer por dos rampas, una desde A hasta C y otra desde C hasta B , el tiempo necesario es de 5.99 s, por lo que el camino más corto no resulta ser el más rápido. Galileo calculó el tiempo de caída que resulta de considerar 2^n rampas que unan los vértices de un polígono regular de 2^{n+2} lados, constatando que cada vez es menor, y concluyó que el camino más rápido para que un cuerpo caiga desde A hasta B es el arco de circunferencia que muestra la figura. En realidad dicho arco es ciertamente más rápido que cualquier concatenación de rampas inscritas en él, pero sucede que no es el camino más rápido. No obstante, analizar este problema requiere un nivel de conocimientos de física y de matemática bastante más sofisticado que el que tenía Galileo.

Otro problema que planteó y resolvió Galileo sobre planos inclinados es el siguiente:

Determinar los puntos desde los cuales, al dejar caer un objeto por un plano inclinado, éste llega a un punto prefijado en un tiempo prefijado.

La solución resulta ser que todos los planos inclinados que empiezan en un punto de una circunferencia y terminan en su base son recorridos en el mismo tiempo, incluyendo el caso de un cuerpo que se deja caer desde el punto más alto H , que no necesita plano inclinado.

Galileo murió en 1642, pero su obra, especialmente sus investigaciones sobre las leyes del movimiento, iba a tener una influencia decisiva en el desarrollo de la ciencia en Occidente.



Los precedentes del cálculo diferencial En 1603 había muerto el matemático francés François Viète. Con su obra se puede decir que el álgebra estaba ya bien consolidada en Occidente y no había dejado de desarrollarse desde entonces. Gracias a las técnicas algebraicas, en las últimas décadas de la vida de Galileo, los matemáticos europeos habían logrado explotar hasta sus últimas consecuencias las técnicas más sofisticadas que habían alcanzado los antiguos griegos, y habían aparecido ya nuevas ideas que superaban conceptualmente las fronteras de la matemática griega. El primer trabajo con ideas esencialmente novedosas fue probablemente una memoria de Fermat que circuló manuscrita durante varias décadas, titulada *Methodus ad disquirendam maximam et minimam et de tangentibus linearum curvarum*, y que tuvo mucha influencia entre los matemáticos de la época.

Como su título indica, Fermat explicaba en ella sus técnicas para encontrar valores máximos y mínimos de expresiones algebraicas, así como para calcular las tangentes de curvas definidas también por expresiones algebraicas. Pero no eran técnicas algebraicas propiamente dichas, sino que Fermat hablaba de cantidades “*adiguales*”, término que tomó prestado de Diofanto, aunque éste lo usaba para referirse a valores aproximadamente iguales en el sentido usual, mientras que para Fermat eran cantidades que se hacen “todo lo iguales que sea posible”. Sus técnicas eran manipulaciones algebraicas que, en un momento dado, se alteraban suprimiendo ciertos términos. Actualmente diríamos que Fermat suprimía cantidades que tienden a 0, pero Fermat no hablaba de nada que remotamente recordara al concepto moderno de límite. Simplemente constataba que sus técnicas proporcionaban los resultados correctos. Era consciente de que no podía justificarlas con rigor, pero terminaba proponiendo un problema a quienes no aprobaran su método, dando a entender que consideraba que no sería posible, o al menos, no sería fácil resolverlo sin él. Se trata del problema de Fermat discutido en la sección 4.5 de [IGE] (encontrar el punto que minimiza la suma de las distancias a tres puntos dados), pero, como vimos allí, fue resuelto en 1640 por Torricelli (un discípulo de Galileo) aplicando técnicas geométricas tradicionales. En efecto, durante varias décadas las nuevas técnicas iban a competir con las tradicionales en la resolución de problemas diversos.

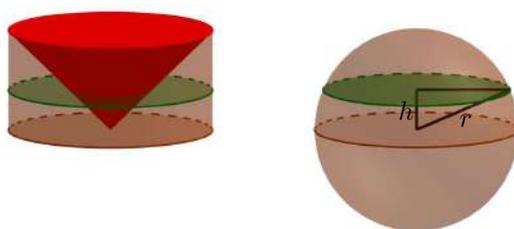
Descartes había desarrollado por su cuenta técnicas similares, pero era consciente de que el rigor de sus razonamientos no alcanzaba lo que se conocía entonces como “rigor geométrico”, así que en su *Géométrie*, publicada en 1637, donde presentó el germen de lo que hoy conocemos como “geometría analítica”, no las mencionó, probablemente porque sentía que desentonaban en un apéndice a su *Discurso del método*, en el que instaba a no aceptar nada que no hubiera sido debidamente justificado por la razón.

Estas técnicas de Fermat y Descartes eran versiones rudimentarias de lo que actualmente es el cálculo de derivadas. Paralelamente se estaban desarrollando los precedentes de lo que ahora conocemos por cálculo integral. Uno de estos precedentes se encuentra en los tratados *Geometria indivisibilibus continuorum noua quandam rationem promota* (Geometría desarrollada por un nuevo método mediante los indivisibles de los continuos), de 1635 y *Exercitationes geometricae sex* (Seis ejercicios geométricos), de 1647, de otro discípulo de Galileo, Bonaven-

tura Cavalieri, en los que expone razonamientos geométricos muy sofisticados para calcular áreas y volúmenes, y enuncia un principio muy simple y práctico a la vez:

Principio de Cavalieri *Si dos figuras están limitadas entre dos mismos planos paralelos, y cualquier plano paralelo intermedio determina en ambas secciones con la misma área, entonces ambas tienen el mismo volumen.*

En realidad este principio había sido usado ya por Arquímedes para calcular el volumen de una esfera. En efecto, Arquímedes observó que si comparamos una semiesfera de radio r con la figura comprendida entre un cilindro y un cono de radio y altura r tal y como muestran las figuras siguientes:



un plano de altura h sobre el plano ecuatorial de la esfera determina en ambas una sección de área $\pi(r^2 - h^2)$, luego el principio de Cavalieri implica que ambas figuras tienen que tener el mismo volumen. Por consiguiente, el volumen de media esfera es

$$V = \pi r^3 - \frac{1}{3}\pi r^3 = \frac{2}{3}\pi r^3,$$

y el volumen de la esfera completa es el doble, es decir:

$$V = \frac{4}{3}\pi r^3.$$

Cavalieri también enunció un principio análogo para calcular áreas, sustituyendo los planos por rectas. Por esa época muchos matemáticos desarrollaron sus propias técnicas de “cuadratura de curvas”, es decir, del cálculo del área limitada por una curva dada, que esencialmente se basaban en el mismo “principio de exhaustión” que había usado Arquímedes y otros geómetras griegos para el cálculo del área de un círculo y otros resultados similares.

Así, por ejemplo, en 1634 Gilles Persone de Roberval aplicó el método de los indivisibles de Cavalieri para calcular el área del arco de cicloide que Galileo no había podido calcular. Concretamente, obtuvo que cada arco de una la cicloide generada por una circunferencia de radio R tiene área $3\pi R^2$. Orgulloso de su hallazgo, se lo comunicó a Descartes, el cual le contestó que era “*un hermoso resultado que no había advertido antes, pero que no causaría ninguna dificultad a ningún geómetra de nivel medio*”. A su vez, Descartes propuso a Roberval y a Fermat calcular las tangentes a una cicloide, cosa que Roberval no consiguió hacer, pero Fermat sí.

En 1641 Torricelli obtuvo un resultado curioso de cuadratura sobre el que él llamó *sólido hiperbólico agudo*, aunque hoy es más conocido como el *cuerno de Gabriel* (en alusión a la trompeta con la que el arcángel Gabriel anunciará el día del Juicio Final). Se trata del sólido limitado por la superficie que resulta del giro de la hipérbola $f(x) = 1/x$ alrededor del eje X desde $x = 1$ en adelante. Torricelli demostró que su volumen es $V = \pi$, pero que el área de su superficie es infinita. He aquí su imagen:



Estamos ante lo que se conoce como la *paradoja del pintor*: el sólido es una “copa” en la que cabe una cantidad finita de pintura, pero ésta no es suficiente para pintar su interior. En palabras de Thomas Hobbes: “*Para verle sentido a esto, no se requiere ser un geómetra o un lógico, sino estar loco*”.

Un ejemplo análogo, pero que puede tratarse de forma mucho más elemental es la “tarta nupcial” formada apilando cilindros de altura 1, el primero de radio 1, el segundo de radio $1/2$, el tercero de radio $1/3$, etc. El volumen total de la tarta será

$$V = \frac{\pi}{1^2} + \frac{\pi}{2^2} + \frac{\pi}{3^2} + \frac{\pi}{4^2} + \frac{\pi}{5^2} + \dots = \frac{\pi^3}{6},$$

mientras que su superficie lateral es

$$A = \frac{2\pi}{1} + \frac{2\pi}{2} + \frac{2\pi}{3} + \frac{2\pi}{4} + \frac{2\pi}{5} + \dots$$

y esta serie es divergente, luego el área es infinita. Así pues, se trata de una tarta que “se puede comer, pero no glasear”.

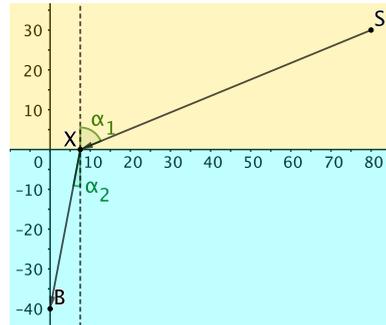
En términos de la paradoja del pintor, si consideramos un recipiente de esta forma lleno de pintura, no resulta extraño que la pintura del primer cilindro sobre para pintar su superficie lateral, y la del segundo sobre para pintar la suya, etc., y que al mismo tiempo, el total de pintura sea finito y la superficie pintada sea infinita. El *quid* está en que cualquier cantidad finita de pintura permite pintar cualquier superficie a condición de que la capa de pintura sea suficientemente fina (y no haya límites a lo fina que puede ser la capa de pintura, lo cual es físicamente imposible).

Por ejemplo, si pensamos que el espesor de la capa de pintura del primer cilindro es de 0.001 mm, debemos ser conscientes de que, a partir de cierto cilindro, ese espesor será mayor que el diámetro, por lo que si consideramos que la superficie lateral del cilindro está pintada con parte de la pintura que contiene, tendremos que admitir que el grosor de la capa sea menor que 0.001, y a medida que pasamos a cilindros superiores tendremos que conformarnos con capas de pintura cada vez más tenues.



Volviendo a Fermat y sus técnicas, he aquí un problema que podía resolverse con ellas:

Un socorrista se encuentra en un punto S a 30 m de la orilla del mar, y ve que un bañista se está ahogando en un punto B situado a 80 m caminando por la costa y a 40 m de profundidad. El socorrista corre hasta la orilla en línea recta hasta cierto punto x y luego se echa al agua. Si corre a 5 m/s y nada a 1 m/s, ¿en qué punto conviene que se eche al agua para llegar al bañista lo antes posible?



Aparte de encontrar la solución, con sus métodos, Fermat podía demostrar un principio general: si v_1 es la velocidad en tierra y v_2 la velocidad en el agua, el punto x que hace el tiempo mínimo hace también que los ángulos α y β mostrados en la figura satisfacen la relación

$$\frac{\text{sen } \alpha_1}{\text{sen } \alpha_2} = \frac{v_1}{v_2}.$$

En realidad Fermat no estaba preocupado por perfeccionar las técnicas de socorrismo, sino que estaba abordando un problema más profundo:

Supongamos que la figura anterior representa la trayectoria de un rayo de luz que viaja por el aire desde el punto S hasta X , y allí entra en otro medio, por ejemplo en el agua. Entonces se refracta, es decir, altera su dirección. Los ángulos α_1 y α_2 que el rayo incidente y el refractado forman con la perpendicular a la superficie que separa los dos medios se llaman *ángulo de incidencia* y *ángulo de refracción*. En 1621, el matemático neerlandés Willebrord Snell van Royen enunció la que hoy se conoce como *ley de Snell*, según la cual la relación entre ambos ángulos es

$$\frac{\text{sen } \alpha_1}{\text{sen } \alpha_2} = \frac{n_2}{n_1},$$

donde n_1, n_2 son unos índices de refracción que dependen de la naturaleza del medio, principalmente de su densidad. En 1658 Fermat dedujo la ley de Snell a partir de un principio general que hoy se conoce como *principio de Fermat*:

El camino seguido por un rayo de luz desde un punto A hasta un punto B es el que requiere el menor tiempo posible para ser recorrido.

Esto literalmente es falso, pues, por ejemplo, si un rayo de luz sale de un punto A , se refleja en un espejo en un punto B y llega a un punto C , habría tardado menos en ir de A a C si hubiera ido en línea recta en lugar de pasando por B . No obstante, Fermat empleó su principio descartando prudentemente estas obviedades y, por otra parte, es posible reformularlo de modo que sea correcto: bajo condiciones físicas adecuadas, el trayecto seguido por un rayo de luz entre dos puntos requiere menos tiempo que cualquier otro trayecto que no se aleje mucho del trayecto real.

Si traducimos la relación que hemos indicado para el problema del socorrista, obtenemos lo que realmente concluyó Fermat, y es que, admitiendo el principio de Fermat, si un rayo de luz se refracta al pasar de un medio en el que se mueve a velocidad v_1 a otro al que se mueve a velocidad v_2 , los ángulos de incidencia y de refracción satisfacen la relación

$$\frac{\sin \alpha_1}{\sin \alpha_2} = \frac{v_1}{v_2} = \frac{c/v_2}{c/v_1},$$

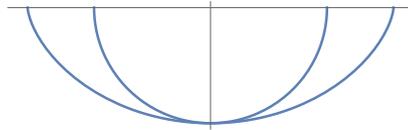
donde c es la velocidad de la luz en el vacío. Así, definiendo los índices de refracción como $n_i = v_i/c$, la fórmula precedente se convierte en la ley de Snell.

Ese mismo año de 1658 Fermat se encontró con un desafío con dos premios publicado por un tal Dettonville, que llamaba a calcular varias características de un arco de cicloide: su longitud, el área que abarca, la superficie que resulta de su giro, etc. El tal Dettonville resultó ser Blaise Pascal, y Fermat y otros matemáticos le presentaron sus soluciones privadamente, sin participar en el desafío. El aporte más novedoso fue el del británico Christofer Wren, quien demostró que la longitud de un arco de cicloide generada por una circunferencia de radio R es $8R$.

Huygens Entre los que enviaron soluciones (correctas) a Pascal se encontraba un matemático y astrónomo neerlandés de 29 años llamado Christiaan Huygens. Doce años atrás, cuando todavía era un estudiante, había entablado correspondencia con Marin Mersenne, que moriría dos años más tarde. En estas cartas tempranas se encuentra una demostración de que la forma de una cadena suspendida de sus extremos no puede ser un arco de parábola, como Galileo había sugerido en su diálogo de 1638, aunque no pudo determinar su forma real.

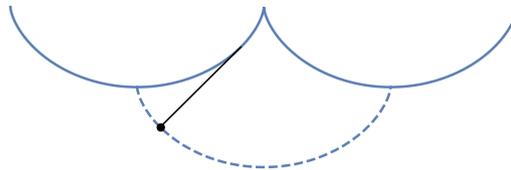
En 1657 Huygens había inventado y patentado el reloj de péndulo. Mersenne le había señalado que la isocronía del péndulo que había descubierto Galileo no se cumple cuando las oscilaciones son grandes, es decir: si dejamos oscilar un péndulo de modo que la desviación máxima respecto a la vertical sea un ángulo pequeño, el periodo de oscilación no se ve modificado a medida que el rozamiento lo va frenando y la amplitud disminuye, pero si las oscilaciones son grandes el periodo ya no es el mismo. Esto no es un inconveniente práctico para el buen funcionamiento de los relojes de péndulo, pues en ellos las oscilaciones son de pequeña amplitud, pero Huygens se planteó el problema de construir un reloj de péndulo que funcionara incluso con oscilaciones de gran amplitud. Para ello abordó primero un problema más simple:

Es fácil ver que el movimiento de un péndulo de longitud r sigue las mismas leyes que el movimiento (sin rozamiento) de una bola sobre una pista semicircular de radio r . Así, la isocronía galileana esquivale a que si dejamos una bola sobre una semicircunferencia a poca distancia de su punto más bajo, empezará a oscilar con un periodo de oscilación que depende del radio de la semicircunferencia, pero no de la altura a la que hemos dejado la bola.² Sin embargo, si la dejamos demasiado alta, el periodo será sustancialmente menor. Huygens se planteó el problema de estirar la semicircunferencia para rebajar su pendiente, de tal forma que el periodo de oscilación sobre la semicircunferencia modificada sea siempre el mismo, independientemente de la altura.



En realidad hay infinitas curvas que cumplen esto, pero sólo hay una que es simétrica respecto del eje vertical, y que está caracterizada por la propiedad de que si se deja caer un objeto por ella, éste llega a la base siempre en el mismo tiempo, independientemente del punto de partida. Huygens llamó *tautócrona* (gr. = el mismo tiempo) a una curva con esta propiedad, y usando técnicas puramente geométricas —Huygens sería toda su vida un matemático de la vieja guardia— demostró que la tautócrona no es sino la cicloide.

La segunda parte del problema era forzar a un péndulo a moverse sobre una cicloide, para lo cual Huygens demostró que si se sujeta un péndulo de longitud L al punto de contacto entre dos arcos de cicloide de longitud $2L$ la trayectoria del péndulo es un arco de cicloide de la misma longitud:



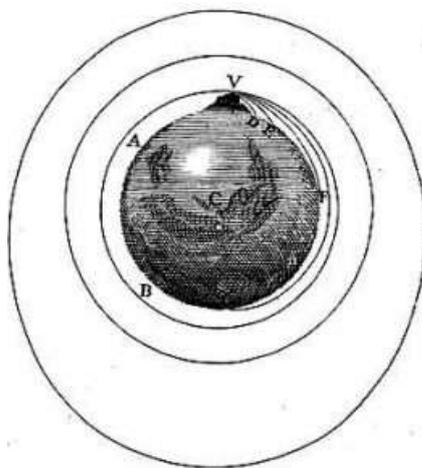
Un péndulo en estas condiciones (mientras el rozamiento sea despreciable) oscila con periodo constante sea cual sea la amplitud de sus oscilaciones.

Newton En 1665 la Universidad de Cambridge tuvo que ser cerrada a causa de una epidemia de peste. Un joven Isaac Newton, recién licenciado a sus 22 años, que había pasado completamente inadvertido como alumno, tuvo que permanecer dos años aislado en su casa, pero en esos dos años revolucionó silenciosamente la física y las matemáticas.

Una de sus reflexiones más trascendentes consistió en plantearse un “experimento mental”. Si lanzamos horizontalmente una bala desde lo alto de una

²En realidad sí que depende de dicha altura, pero, para alturas pequeñas, la dependencia es inapreciable.

montaña, ésta caerá describiendo una trayectoria parabólica. Esto era un hecho sobradamente conocido para cualquier artillero de la época. Cuanto mayor sea la velocidad de salida, más lejos llegará la bala. Con la velocidad suficiente, la bala debería llegar a las antípodas de la montaña, pero, ¿no podría ser lanzada con velocidad suficiente como para que diera una vuelta completa a la Tierra? ¿Y podría ser suficiente como para que llegara con la misma altura y velocidad de partida? Si así fuera, la bala no caería nunca, sino que habría entrado en órbita. Más tarde explicaría esto con la figura siguiente:



En palabras de Newton:

El mismo año [1666] empecé a pensar que la gravedad se extendía a la órbita de la Luna y, tras haber descubierto cómo calcular la fuerza con la que un globo que gira dentro de una esfera presiona la superficie de la esfera, a partir de la regla de Kepler de que los periodos de los planetas estaban en proporción sesquiáltera a sus distancias del centro de sus órbitas, deduje que las fuerzas que mantienen a los planetas en sus órbitas deben ser inversamente proporcionales a los cuadrados de sus distancias del centro en torno al cual giran, y así comparé la Luna en su órbita con la fuerza de la gravedad en la superficie de la Tierra, y descubrí que se correspondían de manera bastante aproximada.

Vamos a interpretar este párrafo:

Imaginemos que hacemos girar una bola dentro de una esfera, de modo que permanece pegada a su superficie por su propio movimiento. Newton conocía el principio de inercia de Galileo, que ahora se conoce también como primera ley de Newton:

Primera ley de Newton *Si un cuerpo está libre de toda influencia externa, o permanece en reposo, o se mueve en línea recta con velocidad constante.*

En virtud de este principio, si la bola sigue una trayectoria circular en lugar de rectilínea, es porque la esfera está ejerciendo una fuerza sobre ella. Newton dice que calculó la fuerza necesaria para que la bola permanezca en su trayectoria circular, pero para ello era necesario precisar cuantitativamente la idea de “fuerza”. El año anterior Newton había desarrollado el aparato matemático necesario para tal fin, y lo denominó “cálculo de fluxiones”, pues llamaba “fluentes” a las magnitudes que varían con el tiempo y “fluxión” de un “fluente” a la velocidad con la que éste varía, definida matemáticamente. Con estos conceptos y el “cálculo de fluxiones” asociado, Newton llegó a un concepto matemático de “fuerza” que capturaba la idea de “influencia que un objeto ejerce sobre otro para modificar su movimiento”, con la cual a su vez pudo formular la que hoy se conoce como segunda ley de Newton:

Segunda ley de Newton *La fuerza total que actúa sobre un cuerpo es igual al producto de su masa por la aceleración que experimenta.*

Aquí la aceleración hay que entenderla como “la velocidad con la que varía la velocidad”. El caso es que Newton pudo hacer un cálculo cuantitativo de la fuerza necesaria para obligar a una bola a seguir una trayectoria circular en lugar de rectilínea. Pero Newton no habla de la fuerza que la esfera ejerce sobre la bola, sino de la fuerza que la bola ejerce sobre la esfera. Aquí está teniendo en cuenta la tercera ley de Newton:

Tercera ley de Newton *Si un cuerpo ejerce una fuerza sobre otro, el segundo ejerce una fuerza sobre el primero de la misma intensidad, pero en sentido contrario.*

En nuestro caso, si la esfera está ejerciendo una fuerza sobre la bola para obligarla a girar (una fuerza dirigida hacia su centro), la bola tiene que estar ejerciendo una fuerza sobre la esfera en dirección opuesta al centro, de modo que es indistinto calcular la intensidad de una o de la otra, pues son iguales.³

A continuación Newton explica que supuso que el movimiento (aproximadamente) circular de la Luna podría explicarse por que la Tierra ejerce una fuerza de atracción sobre la Luna de la intensidad precisa que había calculado para la bola en la esfera. Dicha intensidad depende de la velocidad de rotación, y comparando con la tercera ley de Kepler, llegó a la conclusión de que debía ser inversamente proporcional al cuadrado de la distancia entre el centro de la Tierra y de la Luna.

Más precisamente, Newton obtuvo que la aceleración de Luna no dependía de su masa (igual que sucede con los cuerpos que caen sobre la Tierra, que lo hacen con la misma aceleración sea cual sea su masa), lo cual requería que la

³Hay que advertir que la presentación que estamos haciendo no es históricamente fiel, pues en este punto Newton todavía no entendía correctamente el movimiento circular y no había dado con la formulación correcta de su tercera ley. En realidad Newton planteaba el problema igualando la fuerza centrífuga asociada al movimiento de la Luna con la fuerza centrípeta generada por la gravedad terrestre, un planteamiento por el que un estudiante de física actual podría perfectamente suspender un examen.

fuerza con que la Tierra atrae a la Luna fuera proporcional a la masa de la Luna, y la tercera ley de Newton implicaba a su vez que, dado que la Luna tiene que atraer a la Tierra con la misma fuerza, ésta también tiene que ser proporcional a la masa de la Tierra. En suma, Newton había llegado a la *ley de gravitación universal*:

Ley de gravitación universal *Dos cuerpos cualesquiera se ven atraídos por una fuerza gravitatoria que es directamente proporcional al producto de sus masas e inversamente proporcional al cuadrado de sus distancias.*

Esta ley explicaba el comportamiento de la gravedad en la superficie terrestre y prometía explicar también los movimientos planetarios. Decimos “prometía” porque en realidad Newton no había probado que implicara que los planetas deben moverse en órbitas elípticas. Y no llegó a hacerlo por el momento, porque llegó 1667 y pudo regresar a Cambridge a continuar sus estudios y conseguir un doctorado. Allí mostró algunos de sus resultados sobre el cálculo de fluxiones a su antiguo profesor Isaac Barrow, quien quedó impresionado. Newton se guardó para sí sus reflexiones sobre las leyes del movimiento y la gravedad, y mostró únicamente aplicaciones puramente matemáticas de sus fluxiones, que podían emplearse para calcular áreas, longitudes, tangentes, máximos y mínimos, etc. Sin embargo, no quiso publicar nada porque era consciente de que no podía justificar satisfactoriamente la corrección de sus métodos. Durante los diez años siguientes elaboró al menos tres enfoques diferentes de sus métodos. En 1669 Barrow renunció a su cátedra (para ocupar cargos académicos más elevados) y recomendó a Newton para que la ocupara en su lugar, con lo que Newton consiguió una cátedra a los 26 años.

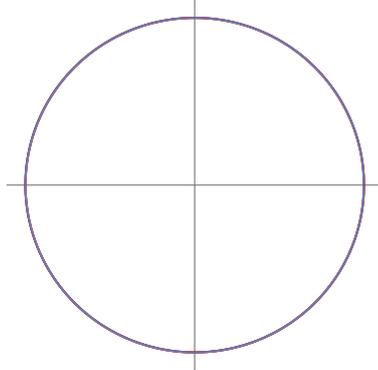
En 1679, a raíz de su correspondencia con Robert Hooke, Newton se interesó de nuevo por la física y en 1684 logró demostrar que las tres leyes de Kepler son consecuencia de la ley de gravitación universal. Finalmente, en 1687 publicó sus *Filosofiae naturalis principia mathematica*, que son probablemente el mayor tratado científico jamás escrito. En él presenta sus tres leyes de la dinámica y la ley de gravitación universal, y a partir de ellas demuestra, entre otras cosas, las leyes de Kepler sobre el movimiento planetario y la generación de las mareas como consecuencia de la gravedad lunar.

Pero Newton seguía sin publicar una exposición sistemática de su cálculo de fluxiones. Aunque había redactado un manuscrito al respecto en 1671 titulado precisamente *El método de los fluxiones*, nunca llegaría a publicarlo. Los planteamientos de los *Principia*, son esencialmente geométricos, con muchas referencias a Euclides y a Apolonio de Perga que simulan “las viejas técnicas geométricas”, pero usa el cálculo de fluxiones “entre bambalinas”, presentando las propiedades que va necesitando sobre las situaciones concretas que trata en cada momento, en las que resultan más plausibles que en enunciados generales.

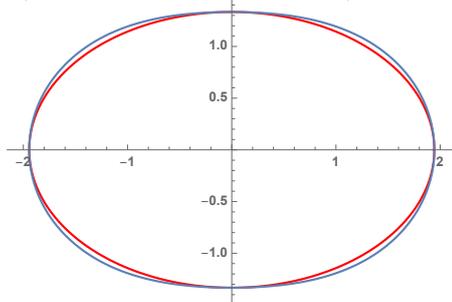
Los *Principia* dieron a Newton fama internacional, aunque los científicos “continentales” preferían la explicación que Descartes había presentado en *El mundo* sobre el movimiento planetario, según el cual los planetas se movían en

remolinos de éter, de modo que no había que suponer que el Sol podía influir sobre la Tierra desde lo lejos (Descartes había fallecido en 1650). De hecho, por esa época, muchos astrónomos seguían defendiendo la teoría geocéntrica. Un caso especial fue el de Giovanni Domenico Cassini, que aceptaba el heliocentrismo, pero no estaba satisfecho con la segunda ley de Kepler. Ahora bien, si se aceptaba la primera, es decir, que los planetas siguen órbitas elípticas, entonces había que aceptar la segunda para que las posiciones de los planetas coincidieran con las observadas. Por ello, en 1693 Cassini se planteó la posibilidad de que las órbitas planetarias no fueran elipses, sino lo que hoy se conoce como óvalos de Cassini.

Mientras una elipse está formada por los puntos del plano tales que la suma de las distancias a dos focos es constante, un *óvalo de Cassini* está formado por los puntos del plano tales que el producto de las distancias a dos focos es constante. Si dejamos de lado a Newton, la hipótesis no era descabellada. La figura siguiente muestra una elipse con la excentricidad de la órbita de Marte y el óvalo de Cassini con los mismos semiejes:



Para apreciar la diferencia consideramos una elipse con excentricidad mayor, por ejemplo, $e = 0.6$ (el óvalo es la curva exterior):



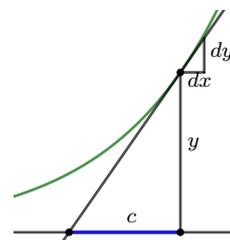
Más concretamente, Cassini se planteó si las órbitas planetarias podrían ser óvalos con el Sol en uno de los focos y de modo que el segundo foco hiciera de ecuante, es decir, que la velocidad (angular) de un planeta fuera uniforme cuando se observa desde el segundo foco, lo cual le parecía más natural que la ley de las áreas. No obstante, años más tarde admitiría que las leyes de Kepler eran correctas.

Leibniz En 1672, durante un viaje a París, Huygens había conocido a un joven diplomático alemán de 26 años llamado Gottfried Wilhelm Leibniz. En 1673 Leibniz ingresó en la Royal Society, a la vez que empezaba a recibir clases de geometría analítica por parte de Huygens, quien le recomendó varios libros. Muy a pesar de su maestro, Leibniz se interesó por las “nuevas técnicas geométricas”, y en 1674 escribió al secretario de la Royal Society comunicándole algunas de sus ideas, pero éste le respondió que Newton tenía métodos generales sobre esos temas. Leibniz estaba obteniendo por su cuenta resultados equivalentes a los del cálculo de fluxiones de Newton, pero para él era más importante lo que Newton había relegado a un segundo plano: sistematizar la “nueva teoría” y, en particular, expresarla en toda su generalidad con la notación más adecuada. Leibniz destacó la fecha del 11 de noviembre de 1675 como uno de los puntos clave de sus investigaciones, cuando introdujo la notación $\int f(x) dx$. En el mismo manuscrito aparece también la fórmula $d(xy) = x dy + y dx$. En 1676 descubrió la fórmula $dx^n = nx^{n-1} dx$, para todo exponente racional n .

Ese año Newton escribió una carta a Leibniz en la que le enumeraba muchos de sus resultados, pero sin demostraciones. Leibniz le respondió inmediatamente, pero la carta había tardado mucho tiempo en llegarle y Newton creyó que se había tomado seis semanas para responderle. Entonces le escribió una segunda carta en tono educado, pero que daba a entender que pensaba que Leibniz le había robado sus métodos. Leibniz le respondió dándole algunos detalles de su propio trabajo, pero Newton señaló —no sin razón— a sus colegas que las técnicas de Leibniz no resolvían ningún problema no resuelto. Por su parte, Leibniz comprendió que debía publicar lo antes posible su trabajo, pero las matemáticas no eran su única ocupación y le llevó un tiempo estar en condiciones de hacerlo.

En 1682 Leibniz instó a Otto Menke a fundar las *Acta Eruditorum*, la primera revista científica alemana, y en su número de octubre de 1684 publicó en ella un artículo titulado: *Noua methodus pro maximis et minimis, itemque tangentibus, quae nec fractas nec irrationales quantitates moratur, et singulari pro illis calculi genus*, (Un nuevo método para máximos y mínimos, así como para tangentes, que no se ve obstaculizado por cantidades fraccionarias o irracionales, y un tipo de cálculo singular para ellas). En él trata el problema de determinar las tangentes a una curva, así como encontrar sus máximos y sus mínimos. También define los conceptos de concavidad, convexidad y punto de inflexión. Las reglas básicas del cálculo con diferenciales son enunciadas sin demostraciones. El trabajo termina así:

Me agradecería añadir como apéndice la solución del problema que Descartes, a propuesta de Beaume, abordó, pero no resolvió: Encontrar una curva de tal naturaleza que la proyección de cualquiera de sus puntos sobre un eje y el corte de con éste de la tangente en el punto determinen un segmento de longitud constante.



Leibniz tradujo el problema a la ecuación diferencial:

$$\frac{dy}{dx} = \frac{y}{c},$$

y la resolvió en pocas líneas, con lo que se convirtió en la primera ecuación diferencial resuelta explícitamente como tal en una publicación matemática.

Leibniz describe el contenido de su artículo como un cálculo que no es geométrico, sino una forma de cálculo simbólico. Esto era una forma elegante de eludir las preguntas delicadas sobre cómo había que entender las “diferenciales” dx , dy que aparecen en sus cálculos (no tienen un significado geométrico, sino que son meros “símbolos”). Pero es interesante ver qué pensaba Leibniz al respecto, por ejemplo, en este párrafo de una carta a John Wallis escrita en 1690:

Es útil considerar las cantidades como infinitamente pequeñas, de modo que, cuando interesa su proporción, pueden omitirse, mejor que verse como 0, cuando se encuentran junto a cantidades que son incomparablemente mayores. Si tenemos $x + dx$, entonces dx se omite. Similarmente, no hay que dejar que $x dx$ y $(dx)^2$ se encuentren uno al lado del otro. Así, si tenemos que diferenciar xy , escribimos $(x + dx)(y + dy) - xy = x dy + y dx + dx dy$. Pero aquí $dx dy$ debe omitirse, porque es incomparablemente menor que $x dy + y dx$. Por lo tanto, en cada caso particular el error es menor que cualquier cantidad finita.

Por ejemplo, veamos cómo Leibniz diferenciaba una raíz cuadrada:

$$\begin{aligned} d\sqrt{y} &= \sqrt{y + dy} - \sqrt{y} = \frac{(\sqrt{y + dy} - \sqrt{y})(\sqrt{y + dy} + \sqrt{y})}{\sqrt{y + dy} + \sqrt{y}} \\ &= \frac{dy}{\sqrt{y + dy} + \sqrt{y}} = \frac{dy}{\sqrt{y} + \sqrt{y}} = \frac{1}{2\sqrt{y}} dy, \end{aligned}$$

donde hemos omitido el dy que acompaña a la cantidad incomparablemente mayor y .

En 1686 Leibniz publicó un segundo artículo en la misma revista, titulado *De geometria recondita et analysi indivisibilium atque infinitorum*, donde aborda el problema de las cuadraturas y demuestra que la integración y la diferenciación son operaciones recíprocas. En él presenta por primera vez el signo \int , al que llama “sumación”. Ese mismo año Leibniz entró en una polémica con los cartesianos sobre si la magnitud adecuada para describir el movimiento de un cuerpo es la cantidad de movimiento considerada por Descartes (el producto de la masa por la velocidad) o bien lo que Leibniz llamaba su *fuerza viva*, que es esencialmente lo que hoy se conoce como energía cinética. En 1687, el año en que Newton publicó sus *Principia*, Leibniz planteó un problema convencido de que los cartesianos no sabrían resolverlo, pues en la solución que él había encontrado era esencial el concepto de “fuerza viva”:

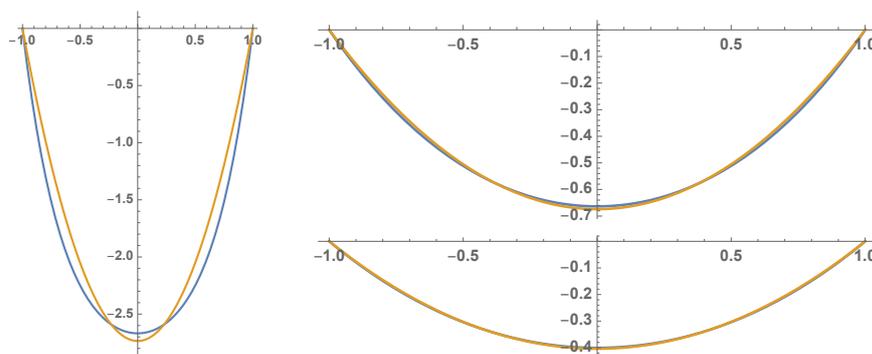
Encontrar la forma de una rampa que haga que los cuerpos que caen por ella se muevan con velocidad vertical constante.

Leibniz llamó *isócrona* a la solución que reclamaba. Sin embargo, unos meses más tarde, Huygens presentó una respuesta geométrica al problema, así que en 1689 Leibniz publicó la suya.

Los Bernoulli Las publicaciones de Leibniz sobre su cálculo diferencial no eran fáciles de digerir y no interesaron a los matemáticos de la época excepto a uno: un matemático suizo llamado Jakob Bernoulli que en 1687, a sus 32 años, había obtenido una plaza en la Universidad de Basilea y que pronto entabló correspondencia con Leibniz para comprender sus técnicas. Al mismo tiempo, estaba enseñando matemáticas a su hermano Johann, de 20 años, y fueron los primeros en asimilar las ideas de Leibniz. Jacob no tardó en resolver el problema de la *isócrona*, y su solución fue publicada en las *Acta eruditorum* en 1690. En su artículo se usó por primera vez la palabra “integral” en su sentido moderno. Ese mismo año fue él quien planteó un nuevo problema:

Encontrar la forma que adquiere una cadena suspendida de sus extremos.

Bernoulli llamó *catenaria* a la curva por la que preguntaba. Según hemos señalado más arriba, Huygens había demostrado a los 17 años que no podía ser una parábola, pero el problema de encontrar la forma correcta seguía abierto. Sin embargo, en 1691, a los 62 años, sí que fue capaz de resolverlo, y su solución fue una de las tres que recibió Jakob, junto con otra de Leibniz y otra de Johann Bernoulli (a las que había que sumar la del propio Jakob). Las figuras muestran las catenarias y las parábolas de longitud $L = 6, 2.5, 2.2$, respectivamente, suspendidas de dos puntos a la misma altura a una distancia $d = 2$ (la catenaria es la que cae menos):



Vemos que para longitudes pequeñas la catenaria es indistinguible de la parábola. Ese mismo año de 1691 Jacob Bernoulli planteó un nuevo problema:

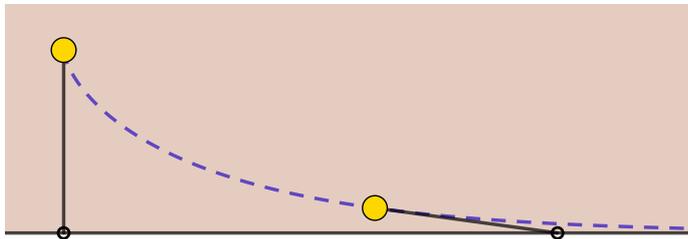
Encontrar la forma que adquiere una varilla elástica cuando se fija un extremo perpendicularmente al suelo y se cuelga de su otro extremo un peso adecuado para que el extremo libre se curve 90° .

Jakob llamó a esta curva la *elástica rectangular*, y tuvo la satisfacción de que nadie supo resolver su problema. En una carta a Leibniz Huygens dijo ese mismo año: *No puedo esperar a ver qué se le ha ocurrido al Sr. Bernoulli el mayor sobre la curvatura de la varilla. Nunca hubiera pensado que se pudiera sacar nada claro o elegante de ahí, así que no lo he intentado.*

En 1693 Leibniz planteó otro problema recordando una conversación que había tenido años atrás en París con Claude Perrault⁴:

El distinguido físico parisino Claude Perrault, igualmente famoso por su trabajo en mecánica y en arquitectura, bien conocido por su edición de Vitrubio, y en vida un importante miembro de la Real Academia Francesa de las Ciencias, me planteó este problema a mí y a muchos otros antes que a mí, admitiendo que no había sido capaz de resolverlo.

Leibniz añadía que, según le había dicho Perrault, ningún matemático había sabido resolver el problema, ni siquiera en Toulouse, que era una forma discreta de decir que Fermat no había sabido resolverlo (Fermat había muerto en 1665). El problema consistía en encontrar la forma de la *tractriz*, es decir, en palabras de Perrault, la curva que describe su reloj de bolsillo puesto sobre una mesa si se tira de su cadena siguiendo el borde:



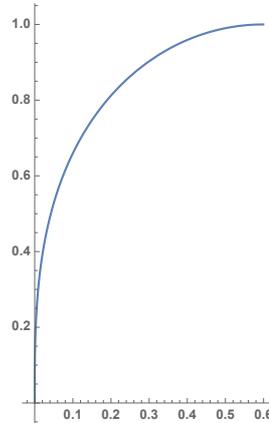
Huygens lo resolvió ese mismo año, pero al parecer Newton ya lo había resuelto en 1676, aunque no había publicado la solución.

En 1694 Jakob Bernoulli publicó su solución al problema de la elástica rectangular, que resultó ser la curva formada por los puntos (x, y) que cumplen

$$x = h \int_0^{y/h} \frac{t^2}{\sqrt{1-t^4}} dt,$$

donde h es la altura de la varilla doblada (la elástica representada más arriba corresponde a $h = 1$). Lo natural hubiera sido expresar la solución en términos

⁴El hermano de Charles, el autor de *La bella durmiente*, *Caperucita Roja*, *El Gato con Botas*, *Cenicienta* o *Pulgarcito*



de la longitud L de la varilla sin doblar, pero ésta viene dada por una integral similar:

$$L = h \int_0^1 \frac{1}{\sqrt{1-t^4}} dt.$$

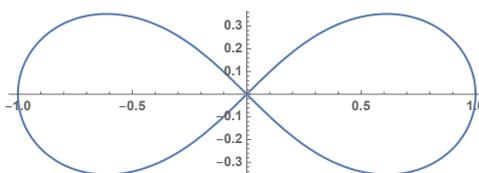
Respecto a estas integrales, Jakob dijo: “*tengo buenas razones para creer que la construcción de nuestra curva no depende ni de la cuadratura ni de la rectificación de ninguna sección cónica*”. Con esto quería expresar que no creía que dichas integrales pudieran resolverse explícitamente (aunque calculó buenas aproximaciones numéricas mediante desarrollos en series de potencias). Jakob consideró que el problema estaría “mejor resuelto” si daba una interpretación geométrica a las integrales, y esto le llevó a buscar una curva “sencilla” cuya longitud viniera dada por la misma integral que la que resulta al calcular la longitud de la elástica rectangular. Así llegó a la curva de ecuación:

$$(x^2 + y^2)^2 = a^2(x^2 - y^2),$$

y probó que su longitud es

$$L = 4a \int_0^1 \frac{1}{\sqrt{1-t^4}} dt.$$

A su vez, el estudio de estas integrales le inspiró para abordar otro problema (el problema de la isócrona paracéntrica) que Leibniz había planteado en 1689 cuando Huygens le estropeó el reto que había planteado a los cartesianos con la isócrona, y ese mismo año publicó un artículo en las *Acta eruditorum* en el que resolvía el problema a la vez que presentaba su nueva curva, a la que describió como una curva con forma de 8 tumbado, o de *lemniscus*, que en latín quiere decir “lazo” o “colgante”, así que pronto fue conocida como “lemniscata”. En general, se llama lemniscata a cualquier curva en forma de 8, por lo que la curva que nos ocupa es, concretamente, la *lemniscata de Bernoulli*:



El caso fue que el artículo de Jakob sobre la elástica también había llevado a su hermano Johann por el mismo camino y también él había descubierto la lemniscata y publicó un artículo sobre ella en las *Acta eruditorum*, con la diferencia de que el artículo de Jakob apareció en septiembre y el de Johann en octubre. A partir de ahí los dos hermanos discutieron sobre quién había sido el descubridor de la lemniscata y nunca se reconciliaron.

La situación era que, en 1694, los únicos matemáticos capaces de resolver los desafíos sobre las aplicaciones del nuevo cálculo diferencial e integral eran el propio Leibniz y los Bernoulli, mientras que Huygens era el único matemático de la

talla necesaria para poder resolverlos con las técnicas geométricas antiguas. Los demás matemáticos de la época, sencillamente no entendían el cálculo diferencial. Naturalmente, Newton era la excepción, que si no había querido participar en los desafíos no había sido, sin duda, por falta de capacidad. Sin embargo, ese año apareció un nuevo aspirante al selecto círculo de los iniciados en el cálculo diferencial:

Johann Bernoulli, a sus 27 años, conoció en París al marqués Guillaume François Antoine de l'Hôpital, quien quedó encantado de haber encontrado a alguien que dominaba el cálculo diferencial, y le ofreció un buen salario a cambio de darle algunas clases, de responder por correspondencia a sus preguntas y de comunicarle sus descubrimientos, con la condición adicional de no revelárselos a nadie más. El joven aceptó estas condiciones que no parecían muy honestas y pronto se vio que no era una mera apariencia. En 1696 el marqués publicó su *Analyse des infiniment petits pour l'intelligence des lignes courbes*, que fue el primer libro de texto sobre cálculo infinitesimal, y era prácticamente idéntico a los apuntes que el propio Bernoulli se había escrito para sí mismo al preparar las clases que le dio a l'Hôpital, si bien éste había corregido algunos errores. Sin embargo, en el libro no aparecía más referencia al verdadero autor que una frase genérica de agradecimiento en el prefacio. En particular, uno de los resultados que Johann Bernoulli le había comunicado al marqués, y que apareció publicado en su libro por primera vez con la misma prueba que puede verse en la correspondencia entre ambos, es hoy conocido como la regla de l'Hôpital. Al parecer, el marqués consideraba que había comprado los derechos intelectuales sobre los resultados de Bernoulli. Éste protestó amargamente en una carta, pero se consideró obligado por el acuerdo que había firmado a no decir nada en vida del marqués.

Ese mismo año de 1696 fue Johann quien planteó por primera vez un desafío a otros matemáticos, que fue publicado en las *Acta eruditorum* con este preámbulo:

Yo, Johann Bernoulli, me dirijo a los matemáticos más brillantes del mundo. No hay nada más atractivo para las personas inteligentes que un problema honesto y desafiante, cuya posible solución le otorgará la fama y permanecerá como un monumento eterno. Siguiendo el ejemplo de Fermat, Pascal, etc., espero ganar la gratitud de toda la comunidad científica al proponer a los mejores matemáticos de nuestro tiempo un problema que pondrá a prueba sus métodos y la fuerza de su intelecto. Si alguien me comunica la solución del problema propuesto, lo declararé públicamente digno de alabanza.

El problema era el siguiente:

Dados dos puntos A y B en un plano vertical, ¿cuál es la curva trazada por un punto que, bajo la acción de la gravedad únicamente, partiendo de A llega a B en el menor tiempo posible?

Más precisamente, se trata de diseñar un tobogán, o incluso un tubo que permita toda clase de cambios de dirección (pero siempre en un mismo plano),

que parta de A y llegue hasta B de modo que si se deja un cuerpo en reposo en A , llegue hasta B en el mínimo tiempo. La curva que resuelve el problema pasó a ser conocida como *braquistócrona*, que en griego significa “la de menor tiempo”.

Bernoulli dijo que le había costado dos semanas dar con la solución. En su desafío había dado un plazo de seis meses, pero nadie respondió y, a petición de Leibniz, lo extendió un año más. Además escribió una carta personalmente a Newton planteándole el problema.

Recientemente Newton había abandonado su vida académica para aceptar el cargo de director de la Casa de la Moneda. Según su sobrina, recibió la carta el 29 de enero de 1697 a las 16:00, cuando regresaba del trabajo, y esa noche a las 4:00 tenía ya la solución. Al día siguiente se la comunicó al presidente de la *Royal Society* y apareció publicada poco después en una revista de la institución, pero de forma anónima porque, según palabras de Newton, *no le gustaba que los matemáticos extranjeros lo agobiaran con cosas de matemáticas*. La respuesta de Newton no daba muchos detalles, pero, en una carta de marzo de 1697, Johann Bernoulli decía que, aunque el autor no había revelado su nombre por un exceso de modestia, los pocos detalles que había dado bastaban para reconocer que era obra de Newton, como *ex ungue leonem* (como se reconoce a un león por su garra).

Finalmente, cinco matemáticos resolvieron el problema: Newton, Leibniz, Jakob Bernoulli, L'Hôpital y Ehrenfried Walter von Tschirnhaus. (Huygens no pudo participar, porque había muerto un año antes de que Johann planteara su desafío.) La braquistócrona resultó ser simplemente una cicloide de modo que la cicloide era a la vez la tautócrona y la braquistócrona, pero la segunda propiedad era más difícil de justificar, pues la tautocronía es una propiedad de la curva en sí, mientras que la braquistocronía requiere comparar la curva con cualquier otra curva posible que una los mismos extremos.

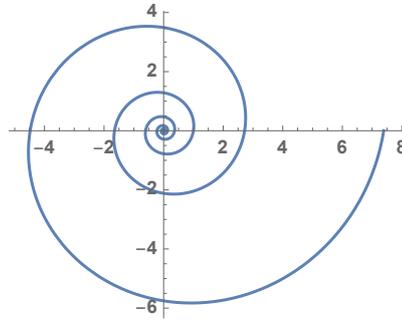
Citamos un curioso resultado que Johann Bernoulli obtuvo en 1697. Leibniz había llegado a descubrir el cálculo integral guiándose en parte por ciertas analogías entre las integrales y las series infinitas, pero la analogía que presenta la primera de las igualdades siguientes entre una integral y una suma parece demasiado hermosa para ser cierta, y sin embargo lo es:

$$\int_0^1 \frac{1}{x^x} dx = \sum_{n=1}^{\infty} \frac{1}{n^n} = \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \frac{1}{4^4} + \dots$$

$$\int_0^1 x^x dx = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^n} = \frac{1}{1^1} - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4} + \dots$$

L'Hôpital murió en 1704, a los 42 años, y Jakob Bernoulli en 1705, a los 50 años. Había dejado dispuesto que en su lápida se grabara una *espiral logarítmica*, a la que había bautizado como *spira mirabilis* (espiral admirable) y que hoy también es conocida como *espiral de Bernoulli*, junto con la frase *Eadem mutata resurgo* (aunque cambiada, resurjo igual). Se trata de una curva descrita por primera vez en 1525 por el pintor Alberto Durero (quien la calificó de “línea

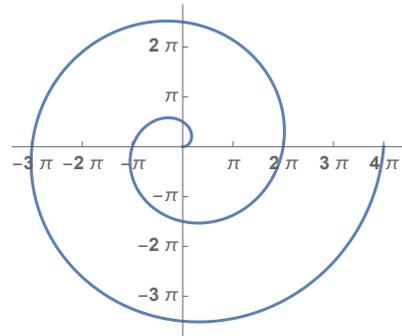
eterna”, y estudiada posteriormente por Descartes en 1638. Se trata de la curva que, en coordenadas polares, viene descrita por la ecuación $\rho = ae^{k\theta}$, es decir, la curva $(x, y) = ae^{k\theta}(\cos \theta, \text{sen } \theta)$.



Bernoulli la estudió a fondo, y observó que, aunque normalmente una homotecia transforma una figura en otra mayor, la espiral logarítmica queda invariante. En efecto, aplicarle una homotecia de razón r equivale a multiplicar sus coordenadas por r , pero expresando $r = e^{k\phi_0}$ vemos que

$$\begin{aligned} r(x, y) &= e^{k\phi_0} ae^{k\theta} (\cos \theta, \text{sen } \theta) = ae^{k(\theta+\phi_0)} (\cos \theta, \text{sen } \theta) \\ &= ae^{k\phi} (\cos(\phi - \phi_0), \text{sen}(\phi - \phi_0)), \end{aligned}$$

y las ecuaciones resultantes son también el resultado de girar la espiral un ángulo ϕ_0 . Esto no sucede, por ejemplo, con una *espiral de Arquímedes*, de ecuación $\rho = k\theta$, en la que las espiras cortan al eje x a intervalos de longitud $2k\pi$, por lo que si aplicamos una homotecia de razón r , la espiral homotética es distinta de la inicial, ya que los intervalos de corte pasan a tener longitud $2kr\pi$.

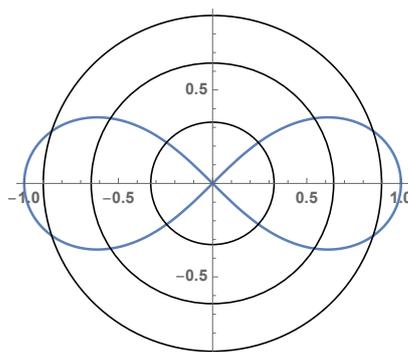


Bernoulli interpretó esta invarianza de la espiral logarítmica como una metáfora de la inmortalidad del alma, y por eso la consideró adecuada para su lápida, pero por error le grabaron una espiral de Arquímedes.

Fagnano Leibniz murió en 1716, a los 70 años, lo que puso fin a una agria polémica sobre quién había descubierto el cálculo infinitesimal. Para entonces

Newton había abandonado la investigación matemática y se empleaba a fondo en perseguir falsificadores de monedas. Johann Bernoulli, a sus 49 años, era entonces el mayor experto mundial en activo sobre cálculo infinitesimal. El tratado de L'Hôpital había contribuido a difundir las nuevas ideas entre la comunidad matemática, pero seguían escaseando los que las dominaban. Uno de los nuevos iniciados de esa época fue el conde Giulio Carlo di Fagnano, que en 1716 había descubierto unos resultados sorprendentes, publicados en 1718.

Los astrónomos estaban aceptando las leyes de Kepler, y el estudio de las órbitas elípticas llevaba de forma natural a cálculos que involucraban longitudes de arcos de elipse, pero sucede que las integrales que permiten calcular dichas longitudes no se pueden calcular. No era que los matemáticos tuvieran argumento alguno que probara la imposibilidad de calcularlas, pero estaba claro que ninguna técnica de integración conocida era aplicable. Y no era el único caso. Lo mismo sucedía con las integrales que determinan la curva elástica, o su longitud, o la longitud de la lemniscata. Parecía que esto impedía cualquier intento de aplicación del cálculo infinitesimal a los problemas relacionados con tales integrales, y ahí es donde Fagnano contradujo las apariencias. Los resultados más espectaculares los obtuvo con la lemniscata.



La figura anterior muestra una lemniscata de semilongitud 1 y tres circunferencias. Fagnano mostró cómo construir esas tres circunferencias con regla y compás, de modo que, junto con el eje horizontal, dividen la lemniscata en 16 arcos de la misma longitud. Más en general, Fagnano mostró cómo dividir una lemniscata con regla y compás en 2^n , $3 \cdot 2^n$ y $5 \cdot 2^n$ arcos iguales, para cualquier número natural n .

Por ejemplo, para la división en 2^n partes iguales, el ingrediente principal de su argumento era el teorema siguiente:

Si $0 < u < 1$ y

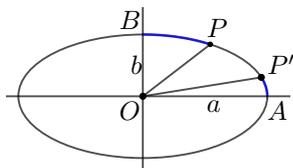
$$v = \sqrt{\frac{-u^2 + \sqrt{2 - 2\sqrt{1 - u^4}}}{1 - \sqrt{1 - u^4}}},$$

entonces

$$\int_0^u \frac{d\rho}{\sqrt{1 - \rho^4}} = 2 \int_0^v \frac{d\rho}{\sqrt{1 - \rho^4}}.$$

Las integrales dan la longitud del arco de lemniscata comprendido entre su centro y los cortes con la circunferencia de radio u y v , de modo que, por ejemplo, para $u = 1$, la circunferencia de radio $v = \sqrt{\sqrt{2} - 1}$ divide un cuadrante de lemniscata en dos arcos iguales, luego divide a la lemniscata en 8 partes iguales, y es la circunferencia intermedia de la figura anterior. Partiendo a su vez del valor $u = \sqrt{\sqrt{2} - 1}$ se obtiene el radio v de la circunferencia menor de la figura, que determina arcos de longitud igual a $1/16$ de la longitud total.

Así, Fagnano no podía calcular las integrales con las que trabajaba, pero eso no le impedía obtener relaciones entre ellas. Sus resultados sobre la elipse eran menos espectaculares, pero igualmente sorprendentes.



Consideremos un punto arbitrario P en una elipse, que podemos representar como $P = (x, y) = (a \cos \phi, b \sin \phi)$, donde el ángulo ϕ no es el ángulo que forma OP con OA , sino el determinado por

$$\frac{b}{a} \tan \phi = \frac{y}{x}.$$

Construimos otro punto $P' = (x', y') = (a \cos \phi', b \sin \phi')$ con la condición de que

$$\tan \phi \tan \phi' = \frac{b}{a}.$$

Explícitamente, sustituyendo en esta condición

$$\tan \phi = \frac{ay}{bx}, \quad \tan \phi' = \frac{ay'}{bx'},$$

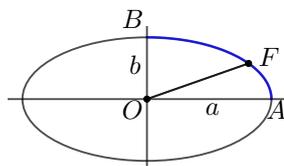
obtenemos que P' es la intersección de la elipse con la recta

$$\frac{y'}{x'} = \frac{b^3 x}{a^3 y}.$$

Es claro que si b/a es constructible con regla y compás, también podemos construir P' a partir de P . Fagnano, aun sin poder calcular las longitudes de los arcos \widehat{BP} y $\widehat{P'A}$, calculó la diferencia de las longitudes:

$$\widehat{BP} - \widehat{P'A} = \epsilon^2 \frac{xx'}{a},$$

donde ϵ es la excentricidad de la elipse.



En particular, si consideramos el que ahora se conoce como *punto de Fagnano* de la elipse, que es el punto $F = F'$ (que es fácil de calcular y de construir con regla y compás), se cumple que

$$\widehat{BF} - \widehat{FA} = b - a.$$

Estos resultados apuntaban a que detrás de aquellas integrales “incalculables” debía de haber alguna teoría matemática no trivial, pero la publicación de Fagnano pasó completamente inadvertida durante décadas.

Euler Johann Bernoulli tuvo tres hijos y una hija. El mayor, Nicolau, murió de apendicitis a los 31 años en 1726, poco después de haber obtenido un puesto de profesor en el departamento de matemáticas y física en la Academia de Ciencias de San Petersburgo. Este pasó a ser ocupado por su hermano Daniel, quien a su vez dejó vacante su plaza en el departamento de fisiología, y recomendó para ocuparla a un amigo suyo de 25 años, Leonhard Euler, a quien su padre, Johann Bernoulli, había estado dando clases de matemáticas los sábados y hasta convencer al padre del muchacho de que el destino de Leonhard no era ser pastor protestante, sino un gran matemático. Y así fue. Euler llegó a San Petersburgo el 17 de mayo de 1727, unos meses después de que Newton falleciera a los 84 años y el mismo día en que murió la emperatriz Catalina I la Grande y su hijo de 12 años se convirtiera en el nuevo zar, Pedro II. Uno de sus tutores era el matemático alemán Christian Goldbach.

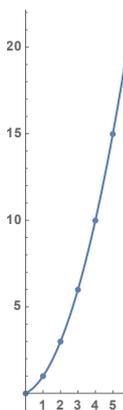
Por esa época estaba de moda el “problema de la interpolación”, consistente en encontrar expresiones analíticas que sobre los números naturales coincidieran con sucesiones dadas. Por ejemplo, la sucesión:

$$1, \quad 1 + 2, \quad , 1 + 2 + 3, \quad 1 + 2 + 3 + 4, \quad \dots$$

que en principio sólo tiene sentido para números naturales, se interpola con la función

$$\frac{x(x+1)}{2}.$$

Es evidente que hay infinitas funciones más que interpolan la misma sucesión, pero se trataba de encontrar “la más natural”, la que se corresponde con la curva que uno dibujaría si no tuviera más intención que unir los puntos dados de la forma más simple posible. La figura muestra la relación entre la sucesión y su interpolación. Goldbach se había interesado en el problema de interpolar la sucesión de los factoriales.



Incapaz de encontrar una expresión interpolante, había consultado a Nicolaus en 1722 y en 1729 planteó el problema a Daniel, quien el 6 de octubre de 1729 le respondió con una carta (Goldbach estaba en Moscú) en la que proponía un producto infinito que interpolaba la sucesión de los factoriales. Daniel le había planteado el problema a Euler, que también había obtenido su propia solución y, animado por Bernoulli, se la comunicó también a Goldbach en una carta del 13 de noviembre. La solución de Euler era:

$$n! = \frac{1^{1-n} \cdot 2^n}{1+n} \cdot \frac{2^{1-n} \cdot 3^n}{2+n} \cdot \frac{3^{1-n} \cdot 4^n}{3+n} \cdot \frac{4^{1-n} \cdot 5^n}{4+n} \dots$$

que se transforma fácilmente en

$$n! = \lim_k \frac{k!(k+1)^n}{(n+1) \cdots (n+k)},$$

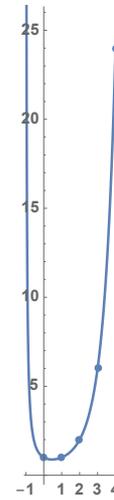
pero en una carta del 8 de enero de 1730 le presentó otra que consideró mucho más satisfactoria:

$$n! = \int_0^1 (-\log t)^n dt.$$

Cambiando la n por una s en la expresión de la derecha obtenemos la función que más adelante Gauss bautizaría como *función factorial*

$$\Pi(s) = \int_0^1 (-\log t)^s dt,$$

cuya gráfica muestra la figura. Esta función no sólo interpola “de forma natural” a la sucesión de los factoriales, sino que cumple en general la ecuación $\Pi(x+1) = (x+1)\Pi(x)$ que define a la sucesión factorial, y aparece hoy en día en contextos muy diversos del análisis matemático y la estadística, si bien en términos de la que Legendre definió como *función gamma*: $\Gamma(x) = \Pi(x-1)$.



Newton había muerto sin haber publicado ningún tratado sobre su método de los fluxiones. En 1734 Euler publicó su tratado *Mechanica*, en el que aplica de forma sistemática el cálculo diferencial en el estudio del movimiento, basándose en leyes generales y no en razonamientos geométricos *ad hoc*, como había hecho Newton. Sin embargo, las técnicas analíticas seguían resultando dudosas — cuando no incomprensibles — para los matemáticos no iniciados. La primera crítica abierta fue publicada en 1734 por el obispo George Berkeley, en un ensayo titulado:

El Analista. Un discurso dirigido a un matemático infiel, en el que se examina si el objeto, los principios y las inferencias del análisis moderno se conciben más claramente o se deducen más evidentemente que los misterios religiosos y las cuestiones de fe.

Berkeley pone el dedo en la llaga en todos los aspectos oscuros del cálculo diferencial. Uno de los pasajes más famosos de su ensayo es éste:

¿Y qué son esos fluxiones? ¿Las velocidades de incrementos evanescentes? ¿Y qué son esos mismos incrementos evanescentes? No son ni cantidades finitas, ni cantidades infinitamente pequeñas, ni tampoco nada. ¿No tendríamos que llamarlos los fantasmas de las cantidades difuntas⁵?

Finalmente, en 1736 se publicó póstumamente *El método de los fluxiones* de Newton, pero no ayudó en nada a refutar las objeciones de Berkeley. El mismo año, Thomas Bayes publicó anónimamente una respuesta a Berkeley titulada

Una introducción a la doctrina de los fluxiones, y una defensa de los matemáticos contra las objeciones del autor de “El analista”, en la medida en que éstas pretenden afectar a sus métodos generales de razonamiento.

Sin embargo, lo que el cálculo infinitesimal necesitaba, no eran defensas dialécticas, sino una fundamentación matemática firme, de la que indudablemente carecía en esa época.

Una buena parte de las matemáticas adquirió su forma actual en los trabajos de Euler. En 1748, un año antes de la muerte de Johann Bernoulli, publicó su *Introductio in analysin infinitorum*, donde introduce el concepto de “función” y afirma que el objeto del análisis matemático es el estudio de las funciones, y en 1755 publicó sus *Institutiones calculi differentialis*. Fue él quien en 1727 introdujo la letra e como nombre para la base de la función exponencial, y en 1734 la notación $f(x)$ para las funciones, y décadas más tarde introduciría la π como nombre del área del círculo unitario, o \sum como signo para los sumatorios, etc., pero estas cuestiones de notación son anecdóticas frente a lo que supuso la obra fundacional de Euler. En sus tratados encontramos por primera vez una exposición lógica, con definiciones, teoremas generales y sus demostraciones correspondientes. Ni las definiciones ni las demostraciones eran en muchos casos rigurosas según los parámetros modernos, pero constituían una base firme tanto para el progreso posterior del cálculo diferencial como para el perfeccionamiento paulatino de sus fundamentos.

Un acontecimiento notable se produjo el 23 de diciembre de 1751, cuando Euler recibió el encargo de revisar las publicaciones de un aspirante a la Academia de Berlín. Se trataba de un libro en dos tomos titulado *Produzioni Matematiche del conte Giulio Carlo di Fagnano, marchese de’ Toschi e di Sant’Onorio*, y había sido publicado el año anterior. En la portada parecía una lemniscata con muchos adornos y con la inscripción *multifariam divisa atque dimensa Deo veritatis gloria* (muchas veces dividida y medida por la gloria del Dios de la verdad).

⁵Berkeley habla de “incrementos evanescentes” por la facilidad con la que, de repente, cuando convenía, se eliminaban de los cálculos. La palabra “*departed*” se usa en inglés habitualmente en el sentido de “difunto”, pero literalmente significa “que se ha marchado”, y Berkeley hace un juego de palabras con estos dos significados. Los infinitésimos son “*departed*” porque se marchan de los cálculos y por eso son como fantasmas de cantidades difuntas.

Así fue como finalmente la obra de Fagnano llamó la atención de Euler, que se sorprendió al ver sus resultados sobre la lemniscata y las elipses. Euler los relacionó con un descubrimiento que él mismo había hecho en 1738, cuando escribió a Johann y a Daniel Bernoulli para comunicarles que, en una elástica rectangular de altura $h = 1$, el producto de su longitud por su desplazamiento es:

$$\int_0^1 \frac{1}{\sqrt{1-t^4}} \cdot \int_0^1 \frac{t^2}{\sqrt{1-t^4}} = \frac{\pi}{4}.$$

Pero el asombro le duró poco, porque al poco tiempo había “cogido el truco” a las técnicas del conde y había generalizado su fórmula de bisección de arcos de lemniscata a un teorema de adición que publicó en 1753. Su enunciado es:

Si $0 < u_1, u_2 < 1$, $u_1 + u_2 < 1$ y

$$v = \frac{u_1 \sqrt{1-u_2^4} + u_2 \sqrt{1-u_1^4}}{1 + u_1^2 u_2^2}.$$

entonces

$$\int_0^{u_1} \frac{d\rho}{\sqrt{1-\rho^4}} + \int_0^{u_2} \frac{d\rho}{\sqrt{1-\rho^4}} = \int_0^v \frac{d\rho}{\sqrt{1-\rho^4}},$$

El resultado correspondiente de Fagnano resulta de hacer $u_1 = u_2 = u$ en el teorema de Euler. Esto significa que, dados dos arcos de lemniscata con un extremo en su centro, es posible construir un tercero cuya longitud sea la suma de las longitudes de los arcos dados.

La prueba de Euler, como las de Fagnano, consistía en un ingenioso cambio de variable que no arrojaba luz alguna sobre por qué la “integral de la lemniscata” satisfacía relaciones tan curiosas. La luz tardaría varias décadas en llegar.

Las integrales elípticas El primer paso lo dio Legendre, que en 1786 empezó a desarrollar una teoría general sobre las integrales de la forma

$$\int R(x, \sqrt{P(x)}) dx,$$

donde $R(x, y)$ es una función racional, es decir, un cociente de polinomios, y $P(x)$ es un polinomio de grado 3 o 4 sin raíces múltiples. Les dio el nombre de *integrales elípticas*, porque entre ellas se encuentra la integral que da la longitud de una elipse. Entre otras cosas, Legendre demostró que toda integral elíptica puede reducirse a tres tipos canónicos de tales integrales, que llamó de *primera*, *segunda* y *tercera especie*, con lo que bastaba estudiar esos tres tipos. Legendre desarrolló una teoría general que le permitió construir tablas a las que podía reducirse el cálculo de cualquier integral elíptica sobre un intervalo concreto.

Cuando Legendre empezó sus investigaciones sobre este tema, Gauss tenía 9 años, pero antes de que las terminara, sobre 1798, Gauss ya había desarrollado su propia teoría basada en una idea revolucionaria.

Gauss se dio cuenta que las manipulaciones farragosas con las que Fagnano y Euler habían demostrado las relaciones que habían encontrado sobre la lemniscata no eran la forma adecuada de estudiar tales integrales. Por ejemplo, si $0 \leq u_1, u_2 \leq 1$ y

$$v = u_1\sqrt{1-u_1^2} + u_2\sqrt{1-u_2^2} \leq 1,$$

se cumple la relación

$$\int_0^{u_1} \frac{dt}{\sqrt{1-t^2}} + \int_0^{u_2} \frac{dt}{\sqrt{1-t^2}} = \int_0^v \frac{dt}{\sqrt{1-t^2}},$$

y puede parecer una relación profunda, similar a la que Euler había obtenido para la lemniscata, pero en realidad es sólo una forma retorcida de expresar la relación

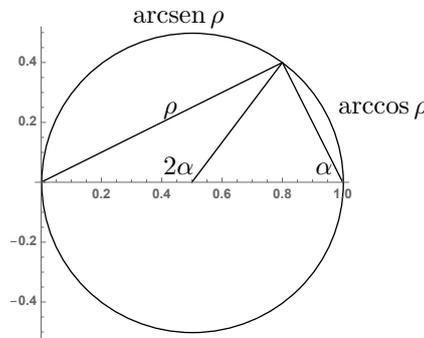
$$\text{sen}(\alpha + \beta) = \text{sen } \alpha \cos \beta + \text{sen } \beta \cos \alpha.$$

En efecto, si llamamos $u_1 = \text{sen } \alpha$, $u_2 = \text{sen } \beta$, para ángulos entre 0 y $\pi/2$, la relación anterior equivale a

$$\arcsen u_1 + \arcsen u_2 = \alpha + \beta = \arcsen\left(u_1\sqrt{1-u_2^2} + u_2\sqrt{1-u_1^2}\right),$$

y los arcos seno se pueden expresar como las integrales indicadas. Gauss demostró que con la fórmula de adición de Euler sucedía algo similar, y descubrió la existencia de una “trigonometría lemniscática”.

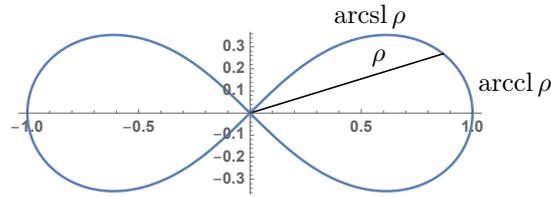
Aunque las funciones trigonométricas se interpretan de forma natural sobre la circunferencia de centro $(0, 0)$ y radio 1, para poner de manifiesto la analogía con el caso de la lemniscata, conviene observar que las funciones arco seno y arco coseno tienen también una interpretación natural sobre la circunferencia de centro $(1/2, 0)$ y radio $1/2$:



Vemos que $\rho = \text{sen } \alpha$, y un arco de amplitud 2α en una circunferencia de radio $1/2$ tiene longitud α , es decir, $\arcsen r$. Similarmente se razona que el arco complementario tiene longitud $\arccos \rho$.

En el caso de la lemniscata, esto llevó a Gauss a definir el arco seno y el arco coseno lemniscáticos como

$$\text{arcsl } \rho = \int_0^\rho \frac{dt}{\sqrt{1-t^4}}, \quad \text{arccl } \rho = \int_\rho^1 \frac{dt}{\sqrt{1-t^4}},$$



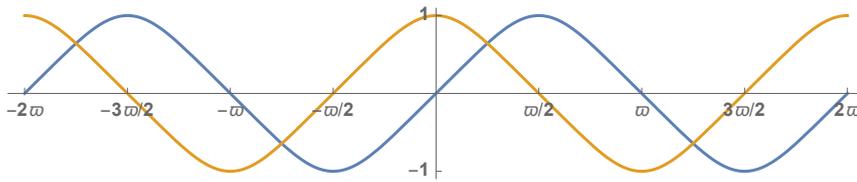
que geoméricamente se interpretan como las longitudes de los arcos de lemniscata determinados por la circunferencia de radio ρ . Así, llamando⁶

$$\varpi = 2 \int_0^1 \frac{dt}{\sqrt{1-t^4}} = 2.622\dots$$

la longitud de la lemniscata completa es 2ϖ , y la de un cuarto de lemniscata es $\varpi/2$, y tenemos la relación

$$\operatorname{arcsl} \rho + \operatorname{arcl} \rho = \frac{\varpi}{2}.$$

Gauss probó que las inversas de estas dos funciones, $\alpha = \operatorname{arcsl} \rho$, $\alpha = \operatorname{arcl} \rho$, a las que llamó el *seno* y el *coseno lemniscático*, $\rho = \operatorname{sl} \alpha$, $\rho = \operatorname{cl} \alpha$, se extienden a funciones periódicas sobre toda la recta real muy similares al seno y el coseno usuales:



En particular satisfacen un teorema de adición similar a la del seno usual, del cual el teorema de Euler es una versión “disfrazada”. Por ejemplo, sustituyendo $\rho = \operatorname{cl} \alpha$ en la relación precedente obtenemos

$$\operatorname{arcsl} \operatorname{cl} \alpha = \frac{\varpi}{2} - \alpha,$$

de donde, aplicando la función sl , llegamos a que $\operatorname{cl} \alpha = \operatorname{sl}(\varpi/2 - \alpha)$, que es la relación análoga a $\cos \alpha = \operatorname{sen}(\pi/2 - \alpha)$.

Por razones que se desconocen, Gauss nunca publicó estos resultados, y la idea de invertir las integrales elípticas para obtener funciones similares a las trigonométricas tuvo que ser redescubierta décadas más tarde. En 1825 Legendre publicó su *Tratado sobre las funciones elípticas* en el que exponía sistemáticamente los resultados que había obtenido, y no fue hasta 1827 cuando dos jóvenes matemáticos de 25 y 23 años, respectivamente, el noruego Niels

⁶En contra de lo que muchos parecen creer, la letra ϖ no es una ω con una raya encima, sino una variante caligráfica de π .

Henrik Abel y el alemán Karl Gustav Jacobi, publicaron con meses de diferencia sendos artículos en los que, a partir de trabajos previos de Legendre, habían llegado a demostrar que las integrales elípticas son inversas de funciones que pueden extenderse, no ya a toda la recta real, sino a todo el plano complejo (salvo a ciertos puntos aislados imaginarios en los que toman el valor infinito) y que no sólo son periódicas como las funciones trigonométricas usuales, sino que son *doblemente periódicas*, en el sentido de que tienen un periodo real y otro imaginario.

Sólo entonces Gauss habló de sus propios resultados, y en una carta a Bessel de 1828 dijo:

No creo probable que vaya a preparar mis investigaciones sobre las funciones trascendentes que he guardado durante muchos años — desde 1798. [...] Herr Abel, como veo, se me ha adelantado y me la librado de esta carga en lo que respecta a una tercera parte de estos temas.

La polémica sobre si Abel o Jacobi habían tomado sus ideas uno del otro se zanjó en 1829 cuando Abel murió de tuberculosis. Ese año Jacobi publicó sus *Fundamenta nova theoriae functionum ellipticarum*, en los que estudió la inversión de la integral elíptica

$$F(\phi | m) = \int_0^\phi \frac{d\theta}{\sqrt{1 - m \operatorname{sen}^2 \theta}} = \int_0^{\operatorname{sen} \phi} \frac{1}{\sqrt{(1 - mt^2)(1 - t^2)}} dt$$

definiendo lo que hoy se conoce como la *amplitud de Jacobi*, que es la función $\phi = \operatorname{am}(u | m)$ que cumple

$$u = \int_0^{\operatorname{am}(u | m)} \frac{d\theta}{\sqrt{1 - m \operatorname{sen}^2 \theta}}.$$

En sí misma, esta amplitud no tiene una interpretación geométrica conocida, pero sucede que en los contextos en los que intervienen integrales elípticas que dependen de un ángulo ϕ , las fórmulas “correctas” se obtienen cuando se expresa $\phi = \operatorname{am}(u | m)$, para un parámetro $0 < m < 1$ adecuado, y trabajando con u en lugar de ϕ . En 1838 Christoph Gudermann introdujo las *funciones elípticas de Jacobi*, dadas por

$$\operatorname{sn}(u | m) = \operatorname{sen} \operatorname{am}(u | m), \quad \operatorname{cn}(u | m) = \operatorname{cos} \operatorname{am}(u | m),$$

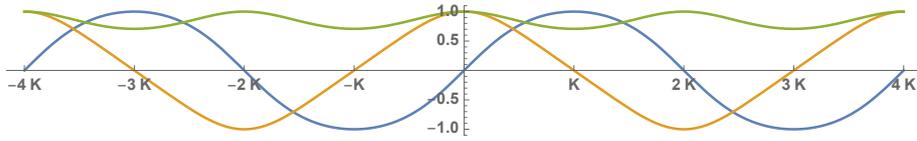
$$\operatorname{dn}(u | m) = \sqrt{1 - m \operatorname{sn}^2(u | m)},$$

de modo que el *seno de la amplitud* sn invierte la segunda de las integrales anteriores:

$$u = \int_0^{\operatorname{sn}(u | m)} \frac{1}{\sqrt{(1 - mt^2)(1 - t^2)}} dt.$$

La tercera función dn es una función auxiliar que aparece, por ejemplo, en la expresión de las derivadas de sn y cn .

He aquí las gráficas para $m = 0.5$:



Las funciones sn y cn tienen periodo $4K$, donde

$$K = K(m) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - m \sin^2 \theta}} = \int_0^1 \frac{1}{\sqrt{(1 - mt^2)(1 - t^2)}} dt,$$

mientras que dn tiene periodo $2K$.

Las funciones elípticas tienen aplicaciones muy diversas. Por ejemplo, la ecuación que rige el movimiento de un péndulo es

$$\theta'' = -\frac{g}{l} \sin \theta,$$

donde g es la intensidad del campo gravitatorio, l la longitud de la cuerda y θ el ángulo que el péndulo forma con la vertical en cada instante. Muchos libros no resuelven esta ecuación diferencial, sino que se limitan a considerar la ecuación

$$\theta'' = -\frac{g}{l} \theta$$

que resulta de aproximar $\sin \theta \approx \theta$, lo cual es aceptable cuando la amplitud máxima de la oscilación θ_0 es pequeña. Entonces la solución es

$$\theta(t) = \theta_0 \sin \left(\sqrt{\frac{g}{l}} t \right),$$

(tomando $t = 0$ como un instante en el que el péndulo está en vertical), y el periodo de las oscilaciones es

$$T = 2\pi \sqrt{\frac{g}{l}}.$$

En particular, T no depende de θ_0 , y esto es la isocronía que advirtió Galileo. Sin embargo, la solución de la ecuación original, sin la aproximación $\sin \theta \approx \theta$ resulta ser

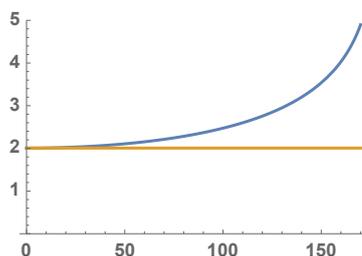
$$\sin \frac{\theta}{2} = k \operatorname{sn} \left(\sqrt{\frac{g}{l}} t \mid k^2 \right),$$

donde $k = \sin \theta_0$, y el periodo es

$$T = 4\sqrt{\frac{g}{l}} K(k^2),$$

que sí que depende de la amplitud θ_0 a través de k .

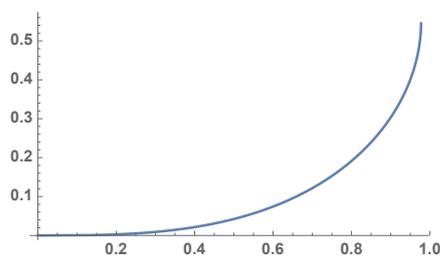
Para un péndulo de longitud $l = 1\text{m}$, el periodo de oscilación dado por la aproximación $T = 2\pi \sqrt{l/g}$ es $T = 2.00709\text{s}$, pero la gráfica siguiente muestra cómo varía el periodo real en función de θ_0 (en grados sexagesimales):



Vemos que para $\theta_0 = 100^\circ$ la diferencia es de casi medio segundo, pero, por ejemplo, la diferencia entre el periodo real y el aproximado para una amplitud de 15° no llega a las 9 milésimas de segundo.

Números complejos Euler fue uno de los precursores del uso de los números complejos en el análisis matemático. A él se debe, por ejemplo, el uso de la notación i para la unidad imaginaria, pero si el uso algebraico de los números imaginarios había resultado “problemático” para los matemáticos renacentistas, las manipulaciones analíticas de Euler con números complejos causaron aún mayor perplejidad a los matemáticos de su época, y sólo fueron debidamente comprendidas a partir de los trabajos de Cauchy sobre funciones de variable compleja, alrededor de 1825.

Una de estas manipulaciones se remonta a un problema planteado en 1694 por Jakob Bernoulli, relacionado con la geometría de las varillas elásticas. Despojando de su interpretación física, lo que planteó Bernoulli era encontrar la forma de una curva cuya curvatura⁷ es directamente proporcional a su longitud.



Mucho después encontró una interpretación más natural, cuando fue redescubierta por el ingeniero estadounidense Arthur Newell Talbot alrededor de 1914, que observó que dicha curva era una de las más adecuadas para conectar tramos de vía, por ejemplo un tramo recto con uno circular. En esa época no era muy relevante porque los trenes no alcanzaban grandes velocidades, pero en un tren que se mueve a gran velocidad (o incluso un coche en una autopista) que la trayectoria no cambie bruscamente de curvatura, sino que ésta aumente

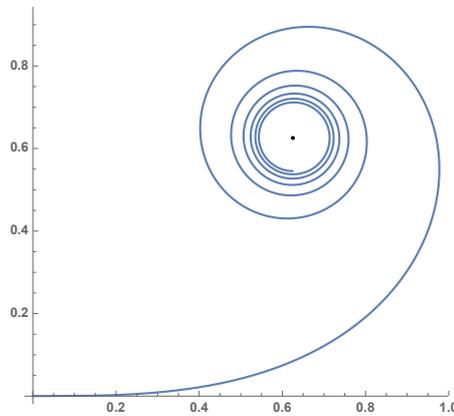
⁷La curvatura de una curva (en el sentido de “lo curvada que está” puede cuantificarse de forma natural por una magnitud llamada precisamente “curvatura”, que definiremos con precisión en su momento, de modo que cuanto mayor es la curvatura, más curvada es la curva. Por ejemplo, una circunferencia tiene curvatura constante igual al inverso de su radio.

o se reduzca gradualmente, evita molestos vaivenes debidos a la inercia, y por ello hoy en día muchos tramos curvos de vías y autopistas (e incluso de pistas de atletismo) tienen la forma de la curva que Bernoulli propuso determinar.

Euler encontró la forma de la curva en 1781, que resultó no admitir una expresión algebraica. Concretamente, probó que (salvo homotecias) se trata de la curva determinada por las ecuaciones

$$x = \int_0^s \cos t^2, dt, \quad y = \int_0^s \sen t^2 dt.$$

Aunque esto ya no tiene interés práctico, si prolongamos indefinidamente la curva obtenemos una espiral, conocida entre otros nombres con el de *espiral de Euler*. Otro nombre más poético se lo dio en 1886 Ernesto Cesàro, a quien la curva le recordó el huso del que la Moira Cloto extrae el hilo que determina el destino de cada ser humano, así que la llamo *clotoide*.



Las integrales $x(s)$, $y(s)$ se conocen como *integrales de Fresnel*, porque en 1818 Augustin Fresnel las usó en el estudio de la difracción de la luz a través de una rendija. Euler se propuso calcular el *polo* alrededor del cual gira la espiral, y llegó a la conclusión de que es el punto de coordenadas $\sqrt{\pi/8}$, es decir, que

$$\int_0^{+\infty} \cos t^2 dt = \int_0^{+\infty} \sen t^2 dt = \sqrt{\frac{\pi}{8}}.$$

El cálculo de estas integrales es uno de los ejemplos de uso polémico de números complejos en un cálculo analítico que presentó Euler. Él consideró una integral más general, pero, particularizando a este caso en concreto, empezó combinando ambas integrales en una sola:

$$\int_0^{+\infty} \cos t^2 dt + i \int_0^{+\infty} \sen t^2 dt = \int_0^{+\infty} (\cos t^2 + i \sen t^2) dt = \int_0^{+\infty} e^{it^2} dt,$$

y comparó esta integral con otra que conocía:

$$\int_0^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

Para relacionar ambas integrales consideró el número complejo

$$\omega = \frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2},$$

que cumple $\omega^2 = -i$, de modo que

$$\int_0^{+\infty} e^{it^2} dt = \int_0^{+\infty} e^{-(\omega t)^2} dt,$$

y a continuación hizo el “cambio de variable” $x = \omega t$, $dx = \omega dt$

$$\int_0^{+\infty} e^{it^2} dt = \int_0^{+\infty} e^{-(\omega t)^2} dt = \omega^{-1} \int_0^{+\infty} e^{-x^2} dx = \left(\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2} \right) \frac{\sqrt{\pi}}{2},$$

de donde, al separar la parte real y la imaginaria, llegamos a que las dos integrales valen $\sqrt{\pi/8}$. Por increíble que pueda parecer, la conclusión es correcta, pero justificarlo no es trivial en absoluto.

En este libro daremos todos los detalles matemáticos de la historia que acabamos de esbozar, daremos definiciones precisas (modernas) de todos los conceptos de los que hemos hablado, demostraremos sus propiedades básicas y resolveremos todos los problemas que hemos planteado y muchos más,⁸ y todo ello reduciendo al mínimo la teoría matemática necesaria. Además de la teoría básica y de algunos de los resultados expuestos en mi *Introducción a la teoría algebraica de números* [ITA], *Introducción a la teoría analítica de números* [ITAn] y la *Introducción a la geometría euclídea* [IGE] (y de unas ligeras generalizaciones a la geometría tridimensional de lo expuesto en este último libro, que presentamos en el apéndice A), nos apoyaremos únicamente en los resultados fundamentales del cálculo diferencial de una y varias variables, reduciendo la topología a la mínima expresión, en particular sin considerar conceptos abstractos como los de “espacio métrico”, “espacio normado”, etc. En el último capítulo presentaremos los resultados más elementales de la teoría de funciones de variable compleja.

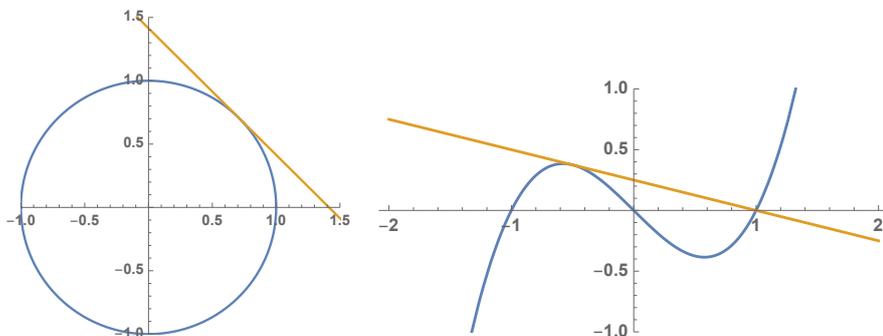
⁸Por no faltar a la verdad, debemos señalar que el único resultado que hemos enunciado y que no demostraremos aquí —aunque lo analizaremos con más detalle— es el principio de Cavalieri, que no es sino un caso particular del teorema de Fubini sobre el cálculo de integrales múltiples.

Capítulo I

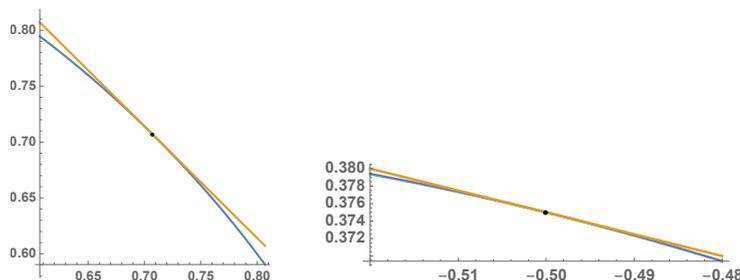
Derivadas

El concepto de derivada de una función tiene esencialmente dos interpretaciones, una “dinámica”, según la cual la derivada mide la velocidad con la que varía una magnitud al variar otra de la cual depende, y otra geométrica, que es la primera que vamos a discutir.

La geometría elemental nos da algunos ejemplos de lo que se llaman “rectas tangentes”. El caso más clásico es el de las rectas tangentes a circunferencias. La definición usual establece que una recta es tangente a una circunferencia si sólo tiene un punto en común con ella, pero esto es engañoso. Se trata de una propiedad que “casualmente” caracteriza la tangencia a una circunferencia, pero en realidad no tiene nada que ver con el fenómeno de la tangencia. Por ejemplo, la figura de la derecha muestra un fragmento de la curva $y = x^3 - x$ y una recta que, a pesar de que corta a la curva en el punto $(1, 0)$, eso no impide que sea tangente en el punto $(-1/2, 3/8)$:



Pero, ¿en qué sentido decimos entonces que la recta es tangente a la curva en $(-1/2, 3/8)$ y secante en $(1, 0)$? La respuesta la obtenemos al ampliar las figuras alrededor del punto de tangencia (véanse las figuras siguientes). En ambos casos vemos que la curva y su recta tangente son casi iguales. Si ampliáramos más aún las figuras, llegaría un momento en que resultarían indistinguibles y no veríamos más que una recta. Ésta es la característica esencial de la noción de “tangencia”:



Una recta es tangente a una curva en un punto si coincide con ella en dicho punto y, cuando se amplía lo suficiente una pequeña porción de ambas alrededor del punto, la recta y la curva resultan indistinguibles.

Nuestro primer problema será dar una formulación matemática precisa a esta idea. Vamos a considerar de momento el caso particular en el que la curva es la gráfica de una función,¹ es decir, que es de la forma $(x, f(x))$, para cierta función $f(x)$ definida alrededor de un punto c , de modo que $(c, f(c))$ es el punto de tangencia. También vamos a suponer que la recta tangente no es vertical, de modo que es también la gráfica de una función de la forma

$$t(x) = f(c) + m(x - c),$$

donde m es la pendiente de la recta. Ésta es la expresión de cualquier recta no vertical que pase por el punto $(c, f(c))$. El problema es determinar cuánto tiene que valer m para que podamos decir que la recta $t(x)$ es tangente a la gráfica de la función f .

La respuesta es simple: La pendiente m de una recta $t(x)$ puede calcularse como

$$m = \frac{t(c+h) - f(c)}{h},$$

para cualquier número $h \neq 0$, pero si sucede que al ampliar una pequeña porción de la gráfica alrededor del punto $(c, f(c))$ la porción ampliada es indistinguible de una recta y tomamos h suficientemente pequeño como para que $c+h$ esté en dicha porción, el valor anterior será indistinguible de

$$\frac{f(c+h) - f(c)}{h},$$

y si queremos que la diferencia sea menor cuanto más ampliamos la gráfica (y, por lo tanto, menor sea la porción ampliada y menor sea h para que $c+h$ esté en ella), lo que necesitamos que se cumpla es que

$$m = \lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h}.$$

¹Notemos que una circunferencia no es la gráfica de una función en este sentido, pero en el ejemplo anterior (correspondiente a la circunferencia de centro $(0,0)$ y radio 1, es decir, la curva de ecuación $x^2 + y^2 = 1$), podemos considerar la función $f(x) = \sqrt{1-x^2}$, para $-1 \leq x \leq 1$, cuya gráfica es media circunferencia.

Vamos a ver que, en efecto, la recta cuya pendiente m viene dada por esta expresión cumple lo requerido, pero antes conviene dar algunas definiciones:

1.1 La derivada de una función

En lo sucesivo la letra \mathbb{K} representará indistintamente al cuerpo \mathbb{R} de los números reales o bien al cuerpo \mathbb{C} de los números complejos.²

Diremos que $A \subset \mathbb{K}$ es un conjunto *abierto* si es un entorno de todos sus puntos, es decir, si para cada $c \in A$ existe un $\delta > 0$ tal que todo $z \in A$ que cumpla $|z - c| < \delta$ cumple $z \in A$.

En otras palabras, A es abierto si cuando contiene a un punto c contiene también a todos los puntos suficientemente próximos o, dicho de otro modo, contiene a todos los puntos de su alrededor.

En el caso en que $\mathbb{K} = \mathbb{R}$, los subconjuntos abiertos más importantes son los intervalos $]a, b[$, $]a, +\infty[$, $] -\infty, b[$ y el propio \mathbb{R} .

Definición 1.1 Sea $f : A \subset \mathbb{K} \rightarrow \mathbb{K}$ una función definida en un conjunto abierto A y sea $c \in A$. Diremos que f es *derivable* en el punto c si existe (y es finito) el límite:

$$\left. \frac{df}{dx} \right|_c = f'(c) = \lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h}.$$

En tal caso, el límite recibe³ el nombre de *derivada* de f en el punto c .

Observemos que, en virtud del teorema [ITAn 3.19], si dos funciones coinciden en un entorno reducido de c en \mathbb{K} y una es derivable en c , lo mismo le sucede a la otra y las derivadas coinciden, es decir, que la existencia y, en su caso, el valor de la derivada de una función en un punto sólo depende de los valores que toma la función alrededor del punto.

En el caso real, si una función f es derivable en un punto c , llamaremos *recta tangente* a la gráfica de f en el punto c a la recta que pasa por $(c, f(c))$ con pendiente $f'(c)$, es decir, la gráfica de la función

$$t(x) = f(c) + f'(c)(x - c).$$

Vamos a probar que la recta tangente así definida cumple la propiedad de tangencia que pretendíamos “capturar” matemáticamente.

Aunque en realidad es irrelevante, pongamos que tomamos una unidad de longitud “apreciable”, por ejemplo 1 cm, y fijemos un $\epsilon > 0$ que sea menor que nuestro umbral de discernimiento visual, por ejemplo, $\epsilon = 0.0001$ (una micra).

²Consideramos aquí los números complejos porque la definición de derivada y sus propiedades básicas valen igualmente para funciones de variable real y compleja, pero de momento sólo discutiremos ejemplos concretos correspondientes a funciones de variable real.

³La notación $f'(c)$ para la derivada se debe a Euler, aunque se conoce como notación de Lagrange, mientras que $df/dx|_c$ es la que empleó Leibniz. También se utiliza la de Newton, que representaba por \dot{y} a la derivada de y .

Por la definición de límite, existe un $\delta > 0$ tal que si se cumple $0 < |h| < \delta$, entonces⁴

$$\left| \frac{f(c+h) - f(c)}{h} - f'(c) \right| < \frac{\epsilon}{5}.$$

Fijemos ahora $\eta < \delta$ y vamos a considerar la ampliación de la gráfica de f en el intervalo $[c - \eta, c + \eta]$ que hace que este intervalo pase a medir 10 cm, lo que equivale a tomar como factor de escala $10/(2\eta) = 5/\eta$. Cada punto $x \in [c - \eta, c + \eta]$, puede expresarse como $x = c + h$, donde $|h| \leq \eta < \delta$. Entonces, la diferencia $|f(x) - t(x)|$ pasará a ser en la figura a escala

$$\begin{aligned} 5 \left| \frac{f(c+h) - f(c) - f'(c)h}{\eta} \right| &< 5 \left| \frac{f(c+h) - f(c) - f'(c)h}{h} \right| \\ &= 5 \left| \frac{f(c+h) - f(c)}{h} - f'(c) \right| < \epsilon, \end{aligned}$$

luego en efecto, la representación a escala de la gráfica de f sobre todo el intervalo $[c - \eta, c + \eta]$ será un rectángulo de 10 cm de lado en el que la diferencia entre $f(x)$ y $t(x)$ será siempre inferior a una micra.

Si queremos una precisión aún mayor sólo tendremos que tomar un δ menor, pero siempre podremos encontrar un intervalo alrededor de c que, al ser ampliado hasta que tenga 10 cm de anchura, la gráfica de f se diferencie de la de la recta tangente en menos del margen ϵ prefijado. Obviamente el haber fijado la unidad de medida en 1 cm o el haber exigido que la gráfica ampliada mida concretamente 10 cm de ancho es irrelevante, pues el razonamiento vale igualmente con otros datos cualesquiera.

Ejemplo Vamos a calcular la derivada de la función $f(x) = x^3 - x$ en el punto $c = -1/2$, y así obtendremos la pendiente de la recta representada en las figuras precedentes. Para ello calculamos, en general,

$$\begin{aligned} f(c+h) - f(c) &= (c+h)^3 - c - h - c^3 + c \\ &= c^3 + 3c^2h + 3ch^2 + h^3 - h - c^3 = 3c^2h + 3ch^2 + h^3 - h, \end{aligned}$$

luego

$$\frac{f(c+h) - f(c)}{h} = 3c^2 + 3ch + h^2 - 1,$$

luego

$$f'(c) = \lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \rightarrow 0} (3c^2 + 3ch + h^2 - 1) = 3c^2 - 1.$$

En particular, para $c = -1/2$ tenemos que $f'(-1/2) = -1/4$, luego la recta tangente a la gráfica de f en $c = -1/2$ es

$$t(x) = f(-1/2) + f'(-1/2)(x + 1/2) = 3/8 - (1/4)(x + 1/2),$$

y esta es, en efecto, la recta que hemos representado en las figuras precedentes. ■

⁴En particular lo siguiente supone que $c + h \in A$, y aquí es donde usamos la hipótesis de que A sea abierto. En caso contrario no tendría por qué cumplirse esto para todo h tal que $0 < |h| < \delta$, sino únicamente para los $h \in A$ que cumplieran esta condición.

En el ejemplo anterior no sólo hemos calculado $f'(-1/2) = -1/4$, sino que hemos comprobado que, en general, $f'(c) = 3c^2 - 1$. Esto es una situación habitual que conviene reflejar en una definición:

Definición 1.2 Diremos que una función $f : A \subset \mathbb{K} \rightarrow \mathbb{K}$ es *derivable* en un conjunto abierto A si existe la derivada $f'(x)$ para todo punto $x \in A$, en cuyo caso la función

$$f' = \frac{df}{dx} : A \rightarrow \mathbb{K}$$

que a cada punto le asigna la derivada de f en dicho punto recibe el nombre de *derivada* de f .

En estos términos, en el ejemplo anterior hemos probado que si $f(x) = x^3 - x$, entonces $f'(x) = 3x^2 - 1$.

En la sección siguiente veremos que las funciones definidas en términos de las operaciones algebraicas básicas y funciones “usuales” (exponenciales, senos, cosenos, logaritmos, etc.) son derivables en todos los puntos en los que están definidas salvo muy pocas excepciones (esencialmente las raíces que se anulan), y las derivadas vienen dadas también por expresiones similares que pueden calcularse aplicando unas sencillas reglas de derivación que no requieren el cálculo explícito de límites. Por ejemplo, veremos (teorema 1.5) que un polinomio

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

es derivable en todo el plano complejo \mathbb{C} y su derivada es

$$f'(z) = n a_n z^{n-1} + (n-1) a_{n-1} z^{n-2} + \cdots + 2 a_2 z + a_1.$$

Derivadas sucesivas A menudo sucede que la derivada de una función derivable es a su vez una función derivable, por lo que puede volverse a derivar, y es habitual que este proceso pueda repetirse incluso indefinidamente, lo que da lugar al concepto de “derivadas sucesivas” de una función.

Así, si $f : A \subset \mathbb{K} \rightarrow \mathbb{K}$ es una función derivable en el abierto A , y su (primera) derivada f' es a su vez derivable, la derivada de f' recibe el nombre de *segunda derivada* de f , y se representa por f'' . Si a su vez ésta es derivable en A , su derivada recibe el nombre de *tercera derivada* de f , y se representa por f''' . Cuando esta notación no resulta práctica se emplea también la notación alternativa $f^{(1)}, f^{(2)}, f^{(3)}, \dots$ y, en general, $f^{(n)}$ representa a la *derivada n -ésima* de la función f (en el supuesto de que exista). Alternativamente, la notación de Leibniz para las derivadas sucesivas es:

$$\frac{df}{dx}, \quad \frac{d^2 f}{dx^2}, \quad \frac{d^3 f}{dx^3}, \quad \cdots, \quad \frac{d^n f}{dx^n}, \quad \cdots$$

Por ejemplo, las derivadas sucesivas del polinomio $f(x) = x^3 - 3x^2 + x - 2$ son

$$f'(x) = 3x^2 - 6x + 1, \quad f''(x) = 6x - 6, \quad f'''(x) = 6,$$

y la cuarta derivada y todas las siguientes son nulas. ■

Tras haber discutido la interpretación geométrica de la derivada de una función como la pendiente de la recta tangente a su gráfica, discutimos ahora la interpretación “dinámica” del concepto de derivada, para lo cual consideramos el ejemplo siguiente:

Ejemplo: La velocidad Veremos más adelante (véase la página 30) que si dejamos caer una piedra desde una altura de 100 m, al cabo de t segundos se encontrará a una altura

$$h(t) = 100 - 4.9t^2 \text{ m.}$$

En general, cuando un cuerpo cae, no lo hace a velocidad constante, sino que su velocidad aumenta a lo largo de la caída. Así, la velocidad inicial de la piedra es de 0 m/s, pero, ¿a qué velocidad está cayendo en el instante $t_0 = 2$ s?

Si partimos de la idea de que la velocidad de un móvil es la distancia que recorre en un segundo, entonces basta observar que $h(2) = 80.4$ m, mientras que $h(3) = 55.9$ m, luego la variación de la altura en un segundo resulta ser de $h(3) - h(2) = -24.5$ m (donde el signo negativo expresa que la altura ha disminuido).

Ahora bien, no es exacto afirmar que en $t_0 = 2$ la piedra caía a una velocidad de 24.5 m/s. Si así fuera, en $t = 2.5$ debería haber descendido $24.5 \cdot 0.5 = 12.25$ m, pero no es el caso: $h(2.5) - h(2) = -11.025$ m, lo que nos lleva a una estimación de la velocidad de:

$$\frac{h(2.5) - h(2)}{0.5} = -\frac{11.025}{0.5} = -22.5 \text{ m/s.}$$

¿Por qué hemos obtenido dos valores diferentes? La respuesta es evidente: el cálculo inicial de -24.5 m/s sólo dice que el espacio recorrido en un segundo es el que habría recorrido un objeto que se hubiera movido *en todo momento* a 24.5 m/s, pero no es el caso de la piedra, que cada vez se mueve más rápidamente, por lo que la velocidad al principio del intervalo $[2, 3]$ es menor que al final. El resultado de 24.5 m/s es una especie de promedio: en realidad la velocidad de la piedra era menor de 24.5 m/s al principio y mayor al final. Por eso al hacer el cálculo considerando el desplazamiento producido en el intervalo $[2, 2.5]$ obtenemos una estimación menor de la velocidad, porque en ese intervalo la velocidad ha tenido menos tiempo para aumentar, luego el aumento paulatino de la velocidad distorsiona menos nuestro intento de medir la velocidad que tenía la piedra justo en el instante $t_0 = 2$.

Pero —siendo consecuentes con esta idea de que, cuanto menos tiempo consideramos, menos varía la velocidad y menos se distorsiona la medición— obtendremos un resultado mejor si consideramos únicamente el espacio recorrido en una centésima de segundo:

$$\frac{h(2.01) - h(2)}{0.01} = -\frac{0.19649}{0.01} = -19.649 \text{ m/s.}$$

En general:

$$\frac{h(2+t) - h(2)}{t} = \frac{100 - 4.9(2+t)^2 - 100 + 4.9 \cdot 2^2}{t} = -19.6 - 4t,$$

y esta expresión muestra que, cuanto menor sea la longitud t del intervalo de tiempo considerado y, por consiguiente, menor sea la distorsión debida a que la velocidad no es constante, la estimación de la velocidad que obtendremos se parecerá más a -19.6 . Por consiguiente, es razonable adoptar como definición que la velocidad del cuerpo en el instante $t_0 = 2$ es exactamente

$$v(2) = h'(2) = \lim_{t \rightarrow 0} \frac{h(2+t) - h(2)}{t} = -19.6 \text{ m/s.}$$

El signo negativo indica, como siempre, que la altura disminuye con el paso del tiempo. Aquí es fundamental entender que si afirmamos que la velocidad de un móvil en un instante t_0 es de 19.6 m/s *no estamos afirmando* que al cabo de un segundo habrá recorrido 19.6 m , sino que *en caso de que mantuviera constante su velocidad*, al cabo de un segundo habría recorrido 19.6 m .

Notemos que esto es coherente con el uso cotidiano de la palabra “velocidad”: si decimos que en un momento dado un coche se está moviendo a 40 km/h , nadie esperará que una hora más tarde el coche haya recorrido 40 km . Perfectamente podría pararse un minuto después, con lo que no habría recorrido más que 666 m , o también podría acelerar hasta los 80 km/h unos segundos después de que hubiéramos medido su velocidad, y entonces —en caso de mantener la nueva velocidad— al cabo de una hora habrá recorrido 80 km en vez de 40 , y nada de ello contradice que en el instante considerado la velocidad fuera de 40 km/h .

Dicho de otro modo: cuando el velocímetro de un coche establece que éste se está moviendo a 40 km/h , no obtiene este resultado dejando que pase una hora y viendo cuánto ha avanzado el coche en ese tiempo, sino que más bien calcula el espacio recorrido en una fracción de segundo y lo divide entre dicha fracción. Si el lapso considerado es suficientemente pequeño, el resultado será una aproximación razonable de la velocidad en un instante dado, pero el valor exacto —si definimos la velocidad como la derivada de la función que nos da el espacio recorrido— lo obtenemos al calcular el límite cuando el intervalo de tiempo considerado tiende a 0 . Esto elimina las distorsiones debidas a que la velocidad puede no haber sido constante en ningún intervalo de tiempo, por pequeño que sea.

Volviendo al ejemplo de la caída libre, observemos que podríamos haber obtenido el mismo resultado aplicando las reglas de derivación para concluir que

$$v(t) = h'(t) = 0 - 4.9 \cdot 2t = -9.8t \text{ m/s,}$$

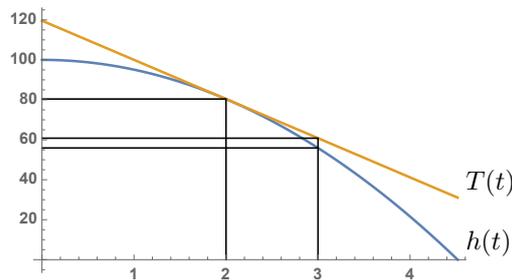
por lo que $v(2) = -19.6 \text{ m/s}$.

Esto nos da una interpretación “dinámica” de la derivada de una función distinta de la interpretación geométrica que le hemos dado anteriormente:

La derivada de una función $f(x)$ en un punto x_0 es la tasa de variación instantánea de $f(x)$ por cada unidad que aumenta la variable x partiendo del valor x_0 , entendiéndose una derivada positiva corresponde a un aumento y una derivada negativa a una disminución.

Aquí es crucial que “tasa de variación instantánea” tiene el sentido preciso que acabamos de discutir y que no es el obvio, en el sentido de que si se cumple, por ejemplo, $f'(7) = 40$, eso no implica que $f(8) = f(7) + 40$, es decir, no implica que al aumentar una unidad el valor $x_0 = 7$ la función vaya a aumentar exactamente 40 unidades, del mismo modo que no cabe esperar que un coche que en $t = 7$ se mueva a 40 km/h vaya a encontrarse exactamente 40 km más lejos en $t = 8$.

Conviene conectar la interpretación de la derivada que estamos discutiendo ahora con su interpretación geométrica. Volviendo al caso de la piedra que cae, la gráfica siguiente muestra la función $h(t)$ y su recta tangente en $t_0 = 2$, que es $T(t) = 80.4 - 19.6(t - 2)$ m/s



La interpretación de esta recta tangente es clara: nos indica la altura a la que se encontraría un objeto que descendiera a una velocidad constante de 19.6 m/s y que en $t_0 = 2$ se encontrara exactamente a la misma altura que la piedra o, equivalentemente, para $t \geq 2$, la función $T(t)$ indica la altura a la que se encontraría la piedra si a partir de ese momento hubiera dejado de acelerar y mantuviera constante su velocidad.

Así, la discrepancia entre $h(3) = 55.9$ m y $T(3) = 60.8$ m se interpreta como que el primer valor es la altura real de la piedra en $t = 3$, mientras que el segundo es la altura a la que se encontraría la piedra en $t = 3$ si no hubiera ido aumentando su velocidad paulatinamente. ■

Curvas paramétricas Una curva no tiene por qué ser expresada como la gráfica de una función, sino que una forma mucho más versátil de presentar una curva es en forma paramétrica, es decir, mediante una función $r : I \rightarrow \mathbb{R}^3$, donde I es un intervalo en \mathbb{R} (que por conveniencia tomaremos abierto). Más explícitamente, una función x en estas condiciones es de la forma⁵

$$r(t) = (x_1(t), x_2(t), x_3(t)) = (x(t), y(t), z(t)).$$

Por ejemplo, una elipse puede parametrizarse mediante la función

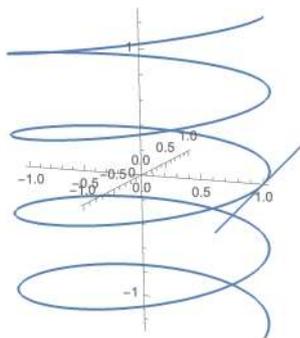
$$r(t) = (a \cos t, b \sin t).$$

⁵Todo cuando vamos a decir se aplica igualmente a curvas en \mathbb{R}^2 sin más que eliminar la tercera componente.

Si pensamos en t como una variable temporal, entonces $x(t)$ es la posición en el instante t de la “punta del lápiz” que dibuja la curva, o la posición de un objeto puntual que recorre la curva. Un ejemplo de curva tridimensional es la *hélice*

$$r(t) = (r \cos t, r \sin t, st),$$

que gira alrededor del eje Z a la vez que asciende.



Definición 1.3 Una curva paramétrica $r :]a, b[\rightarrow \mathbb{R}^3$ es *derivable* en un punto $t_0 \in]a, b[$ si sus funciones coordenadas $x_i(t)$ son derivables en t_0 y, en tal caso definimos su *derivada* como

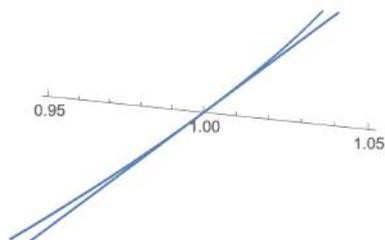
$$r'(t_0) = \left. \frac{dr}{dt} \right|_{t_0} = (x'(t_0), y'(t_0), z'(t_0)).$$

Más aún, si $r'(t_0) \neq 0$, definimos la *recta tangente* a la curva en el punto $r(t_0)$ como la recta dada por

$$T(t) = r(t_0) + r'(t_0)(t - t_0).$$

(Las derivadas sucesivas se definen de forma obvia.)

Vamos a probar que esta recta es realmente tangente a la curva en el sentido que hemos introducido informalmente en la introducción a este capítulo de que se confunde con ella al ampliar suficientemente la curva. Por ejemplo, la figura anterior muestra la recta correspondiente a $t_0 = 0$ de la hélice $r(t) = (\cos t, \sin t, 0.1t)$. Aquí vemos una ampliación de la curva y la recta alrededor del punto de contacto:



Aunque es irrelevante, pongamos que la unidad de longitud es 1 cm y que $\epsilon = 0.0001$ es una micra. Por la definición de derivada, existe un $\delta > 0$ tal que si $0 < |h| < \delta$, entonces

$$\left| \frac{x_i(t_0 + h) - x_i(t_0) - x'_i(t_0)h}{h} \right| < \frac{\|r'(t_0)\|}{5\sqrt{3}} \epsilon,$$

donde $x_1(t), x_2(t), x_3(t)$ son las tres funciones coordenadas de $r(t)$. Fijemos cualquier $\eta < \delta$ y vamos a ampliar los fragmentos de la curva y de la recta correspondientes a los valores del parámetro en el intervalo $[t_0 - \eta, t_0 + \eta]$ con un factor de escala de $5/(\eta\|r'(t_0)\|)$. El segmento de la recta es el que tiene por extremos los puntos

$$T(t_0 \pm \eta) = r(t_0) \pm r'(t_0)\eta,$$

cuya longitud es

$$\|T(t_0 + \eta) - T(t_0 - \eta)\| = \|2r'(t_0)\eta\| = 2\eta\|r'(t_0)\|,$$

y al aplicarle el factor de escala se convierte en un segmento de longitud 10 cm. Vamos a ver que cada punto de este segmento se diferencia en menos de una micra del punto de la curva correspondiente al mismo parámetro. En efecto, para un $t = t_0 + h$, con $0 < |h| \leq \eta$, la diferencia (ampliada con el factor de escala que hemos fijado) es:

$$\begin{aligned} \frac{5}{\eta\|r'(t_0)\|} \|r(t) - T(t)\| &= \frac{5}{\eta\|r'(t_0)\|} \|r(t_0 + h) - r(t_0) - r'(t_0)h\| \\ &\leq \frac{5}{\|r'(t_0)\|} \left\| \frac{r(t_0 + h) - r(t_0) - r'(t_0)h}{h} \right\| \\ &= \frac{5}{\|r'(t_0)\|} \sqrt{\sum_{i=1}^3 \left(\frac{x_i(t_0 + h) - x_i(t_0) - x'_i(t_0)h}{h} \right)^2} \\ &< \frac{5}{\|r'(t_0)\|} \sqrt{\sum_{i=1}^3 \left(\frac{\|r'(t_0)\|\epsilon}{5\sqrt{3}} \right)^2} = \frac{5}{\|r'(t_0)\|} \sqrt{3 \left(\frac{\|r'(t_0)\|\epsilon}{5\sqrt{3}} \right)^2} = \epsilon. \end{aligned}$$

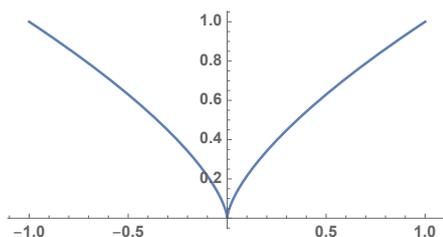
Así pues, el fragmento ampliado de la curva se diferencia en menos de una micra de un segmento de recta de 10 cm de longitud. Obviamente, las medidas concretas que hemos tomado son irrelevantes, y el mismo razonamiento se aplica a otras cualesquiera. ■

Notemos que la definición de recta tangente a una curva paramétrica incluye como caso particular a la definición de recta tangente a la gráfica de una función, pues si $f(x)$ es una función derivable en un punto x_0 , su gráfica puede parametrizarse como $r(x) = (x, f(x))$, y su recta tangente según la definición que acabamos de dar es

$$T(x) = (x_0, f(x_0)) + (1, f'(x_0))(x - x_0) = (x, f(x_0) + f'(x_0)(x - x_0)),$$

que es la gráfica de la función $t(x) = f(x_0) + f'(x_0)(x - x_0)$, es decir, la recta tangente en el sentido que hemos definido para gráficas.

Nota Si una curva $r(t)$ es derivable en t_0 pero $r'(t_0) = 0$, entonces no tenemos definida la recta tangente, pero es que ésta no tiene por qué existir. Por ejemplo, la figura siguiente muestra la curva $r(t) = (t^3, t^2)$, que cumple $r'(0) = (0, 0)$ y, en efecto, la gráfica muestra que no tiene recta tangente en $r(0) = (0, 0)$, pues por mucho que amplíemos la gráfica nunca dejaremos de ver en ella un “pico”, luego la gráfica nunca se confundirá con la de una recta.



1.2 Cálculo de derivadas

Veamos ahora que, tal y como hemos anticipado en la sección precedente, el cálculo práctico de la derivada de una función “usual” se reduce a la aplicación de unas sencillas “reglas de derivación”.

Para empezar conviene observar lo siguiente:

Teorema 1.4 Si una función $f : A \subset \mathbb{K} \rightarrow \mathbb{K}$ es derivable en un punto $c \in A$, entonces f es continua en c .

DEMOSTRACIÓN: Se cumple que

$$\lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} = f'(c),$$

luego, usando que $\lim_{h \rightarrow 0} h = 0$ y que el límite de un producto es el producto de los límites:

$$\lim_{h \rightarrow 0} f(c+h) - f(c) = 0,$$

luego

$$\lim_{h \rightarrow 0} f(c+h) = f(c).$$

Pero esto implica que

$$\lim_{x \rightarrow c} f(x) = f(c).$$

En efecto, dado $\epsilon > 0$, existe un $\delta > 0$ tal que si $0 < |h| < \delta$, entonces $|f(c+h) - f(c)| < \epsilon$, luego si $0 < |x - c| < \delta$, llamando $h = x - c$, tenemos que $|h| < \delta$, luego $|f(x) - f(c)| < \epsilon$, y esto significa que el límite de $f(x)$ es $f(c)$. A su vez, esto equivale a la continuidad de f en c . ■

El teorema siguiente basta para derivar cualquier cociente de polinomios:

Teorema 1.5 (Reglas básicas de derivación):

- a) Toda función constante $f(x) = c$ es derivable y $f'(x) = 0$.
- b) La función $f(x) = x$ es derivable, y $f'(x) = 1$.
- c) Si $f, g : A \rightarrow \mathbb{K}$ son derivables en un punto x , entonces $f + g$ y fg también son derivables en x y

$$(f + g)'(x) = f'(x) + g'(x), \quad (fg)'(x) = f'(x)g(x) + f(x)g'(x).$$

En particular, si g es derivable en x , también lo es cg y $(cg)'(x) = cg'(x)$. Si además $g(x) \neq 0$, también es derivable f/g y

$$(f/g)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

- d) Si $f(x) = x^m$, donde m es un número entero, entonces $f'(x) = mx^{m-1}$.

DEMOSTRACIÓN: a) La derivada de una función constante es

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{c - c}{h} = \lim_{h \rightarrow 0} 0 = 0.$$

b) La derivada de $f(x) = x$ es

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{x+h-x}{h} = \lim_{h \rightarrow 0} 1 = 1.$$

c) Claramente:

$$\begin{aligned} (f + g)'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) + g(x+h) - f(x) - g(x)}{h} \\ &= \lim_{h \rightarrow 0} \left(\frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h} \right) = f'(x) + g'(x). \\ (fg)'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} g(x+h) + f(x) \frac{g(x+h) - g(x)}{h} = f'(x)g(x) + f(x)g'(x), \end{aligned}$$

donde hemos usado la continuidad de g . Si además $g(x) \neq 0$, la continuidad de g implica a su vez que, si h es suficientemente pequeño, se cumple que $g(x+h) \neq 0$. Teniendo esto en cuenta:

$$(f/g)'(x) = \lim_{h \rightarrow 0} \frac{f(x+h)/g(x+h) - f(x)/g(x)}{h}$$

$$\begin{aligned}
&= \lim_{h \rightarrow 0} \frac{f(x+h)g(x) - f(x)g(x+h)}{hg(x+h)g(x)} \\
&= \lim_{h \rightarrow 0} \frac{f(x+h)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(x+h)}{hg(x+h)g(x)} \\
&= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{hg(x+h)} - \frac{f(x)(g(x+h) - g(x))}{hg(x+h)g(x)} \\
&= \frac{f'(x)}{g(x)} - \frac{f(x)g'(x)}{g(x)^2} = \frac{f(x)g(x) - f(x)g'(x)}{g(x)^2}.
\end{aligned}$$

d) Si $m = 0, 1$ es consecuencia de a) y b). Supuesto cierto para un número natural m , entonces $f(x) = x^{m+1} = g(x)x$, donde $g(x) = x^m$, luego por c) y la hipótesis de inducción,

$$f'(x) = mx^{m-1}x + x^m \cdot 1 = mx^m + x^m = (m+1)x^m.$$

Ahora, si $f(x) = x^m$, con $m < 0$, tenemos que $f(x) = 1/x^{-m}$, luego, por la regla del cociente,

$$f'(x) = \frac{0 \cdot x^{-m} - 1 \cdot (-m)x^{-m-1}}{x^{-2m}} = mx^{m-1}. \quad \blacksquare$$

La tabla siguiente contiene las derivadas de las funciones más importantes:

Tabla 1.1: Derivadas de las funciones elementales

$f(x)$	$f'(x)$	$f(x)$	$f'(x)$
e^x	e^x	$\operatorname{sen} x$	$\cos x$
$\log x$	$1/x$	$\cos x$	$-\operatorname{sen} x$
a^x	$a^x \log a$	$\tan x$	$1/\cos^2 x$
x^a	ax^{a-1}	$\operatorname{senh} x$	$\cosh x$
$\sqrt[n]{x}$	$\frac{1}{n\sqrt[n]{x^{n-1}}}$	$\cosh x$	$\operatorname{senh} x$
		$\tanh x$	$1/\cosh^2 x$
$\operatorname{arcsen} x$	$\frac{1}{\sqrt{1-x^2}}$	$\operatorname{arsenh} x$	$\frac{1}{\sqrt{x^2+1}}$
$\operatorname{arccos} x$	$-\frac{1}{\sqrt{1-x^2}}$	$\operatorname{arcosh} x$	$\frac{1}{\sqrt{x^2-1}}$
$\operatorname{arctan} x$	$\frac{1}{1+x^2}$	$\operatorname{artanh} x$	$\frac{1}{1-x^2}$

Algunas de ellas son fáciles de justificar. Por ejemplo, la derivada de e^x es

$$\lim_{h \rightarrow 0} \frac{e^{x+h} - e^x}{h} = \lim_{h \rightarrow 0} e^x \frac{e^h - 1}{h} = e^x,$$

donde hemos usado la fórmula (4.4) de [ITAn].

Igualmente, teniendo en cuenta las observaciones tras [ITAn 4.5], la derivada de $\log x$ es

$$\lim_{h \rightarrow 0} \frac{\log(x+h) - \log x}{h} = \lim_{h \rightarrow 0} \frac{1}{x} \frac{\log(1+h/x)}{h/x} = \frac{1}{x}.$$

Para derivar $\sin x$ calculamos:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin x}{h} &= \lim_{h \rightarrow 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\ &= \lim_{h \rightarrow 0} \sin x \frac{\cos h - 1}{h} + \cos x \frac{\sin h}{h} = \cos x, \end{aligned}$$

donde usamos que

$$\lim_{h \rightarrow 0} \frac{\cos h - 1}{h} = 0, \quad \lim_{h \rightarrow 0} \frac{\sin h}{h} = 1.$$

Para el coseno podemos razonar de forma similar, o bien aplicamos la regla de la cadena (el teorema siguiente) a la relación $\cos x = \sin(\pi/2 - x)$, de modo que la derivada del coseno es

$$-\cos(\pi/2 - x) = -\sin x.$$

Para derivar $\tan x = \sin x / \cos x$ usamos la regla del cociente:

$$\frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}$$

Conviene observar que de la relación $\sin^2 x + \cos^2 x = 1$ se sigue que

$$\frac{1}{\cos^2 x} = 1 + \tan^2 x,$$

luego ésta es una expresión alternativa para la derivada de la tangente.

Sin embargo, con lo visto hasta ahora no tenemos ninguna forma sencilla de calcular la derivada de una función como $\sin x^2$. Para derivar funciones como ésta necesitamos la regla de la cadena:

Teorema 1.6 (Regla de la cadena) *Sea $y(x)$ una función derivable en un punto x_0 y sea $z(y)$ una función derivable en $y_0 = y(x_0)$. Entonces la función compuesta $z(x) = z(y(x))$ es derivable en x_0 y $z'(x_0) = z'(y_0)y'(x_0)$ o, equivalentemente:*

$$\left. \frac{dz}{dx} \right|_{x_0} = \left. \frac{dz}{dy} \right|_{y_0} \left. \frac{dy}{dx} \right|_{x_0}.$$

DEMOSTRACIÓN: Aquí hay que entender que la función $y(x)$ está definida en un abierto $A \subset \mathbb{K}$ que contiene a x_0 , y que toma valores en un abierto sobre el que está definida la función $z(y)$, con lo que la función compuesta $z(x)$ está definida en A . La idea fundamental de la prueba está contenida en el argumento siguiente que, en realidad, tiene un fallo, pero que es fácilmente subsanable:

Si $h \neq 0$ es un número suficientemente pequeño como para que $x + h \in A$, llamamos $h' = y(x_0 + h) - y_0$, de modo que

$$\begin{aligned} \frac{z(x_0 + h) - z(x_0)}{h} &= \frac{z(y(x_0 + h)) - z(y_0)}{h} = \\ &= \frac{z(y_0 + h') - z(y_0)}{h'} \cdot \frac{y(x_0 + h) - y(x_0)}{h}. \end{aligned}$$

Como $y(x)$ es derivable en x_0 , en particular es continua, luego si $h \rightarrow 0$, se cumple también que $h' = y(x_0 + h) - y_0 \rightarrow 0$, luego la derivabilidad de $y(x)$ y $z(y)$ nos da que

$$\begin{aligned} z'(x_0) &= \lim_{h \rightarrow 0} \frac{z(x_0 + h) - z(x_0)}{h} = \\ &= \lim_{h' \rightarrow 0} \frac{z(y_0 + h') - z(y_0)}{h'} \cdot \lim_{h \rightarrow 0} \frac{y(x_0 + h) - y(x_0)}{h} = z'(y_0) \cdot y'(x_0). \end{aligned}$$

El fallo principal que tiene el razonamiento anterior es que no tiene en cuenta que puede haber valores de h para los que $h' = 0$, con lo que los cocientes que hemos planteado pueden no tener sentido. Para evitar este problema es preferible trabajar con una sucesión $\{h_n\}_{n=0}^{\infty}$ de números no nulos. Si probamos que existe el límite

$$\lim_n \frac{z(x_0 + h_n) - z(x_0)}{h_n} = z'(y_0) \cdot y'(x_0)$$

el teorema estará demostrado. Para ello distinguimos varios casos:

1) Para todo n suficientemente grande, $h'_n = y(x_0 + h_n) - y_0 \neq 0$. Entonces, por la continuidad de $y(x)$ en x_0 tenemos que $\lim_n h'_n = 0$, luego el cálculo anterior cambiando h por h_n nos da que, ciertamente,

$$\begin{aligned} \lim_n \frac{z(x_0 + h_n) - z(x_0)}{h_n} &= \lim_n \frac{z(y_0 + h'_n) - z(y_0)}{h'_n} \cdot \lim_n \frac{y(x_0 + h_n) - y(x_0)}{h_n} \\ &= z'(y_0) \cdot y'(x_0). \end{aligned}$$

2) Si para todo n suficientemente grande $y(x_0 + h_n) = y_0$, entonces

$$\frac{z(x_0 + h_n) - z(x_0)}{h_n} = \frac{z(y(x_0 + h_n)) - z(y(x_0))}{h_n} = 0,$$

y, por otra parte,

$$y'(x_0) = \lim_{h \rightarrow 0} \frac{y(x_0 + h) - y(x_0)}{h} = \lim_n \frac{y(x_0 + h_n) - y(x_0)}{h_n} = \lim_n 0 = 0,$$

luego

$$\lim_n \frac{z(x_0 + h_n) - z(x_0)}{h_n} = 0 = z'(y_0)y'(x_0).$$

3) Si hay valores de n arbitrariamente grandes para los que $h_n \neq 0$ y valores para los que $h_n = 0$, descomponemos la sucesión $\{h_n\}_{n=0}^{\infty}$ en dos subsucesiones y, por los dos casos anteriores, ambas cumplen la igualdad

$$\lim_n \frac{z(x_0 + h_n) - z(x_0)}{h_n} = z'(y_0)y'(x_0),$$

luego lo mismo vale para la sucesión completa. ■

Ahora podemos justificar fácilmente las reglas de derivación de algunas funciones más de la tabla 1.1:

Por ejemplo, sin más que recordar que (para $a > 0$) $a^x = e^{x \log a}$, la regla de la cadena nos da que la derivada de a^x es $e^{x \log a} \log a = a^x \log a$. Similarmente, derivando $x^a = e^{a \log x}$ (para $x > 0$) por la regla de la cadena queda

$$e^{a \log x} \frac{a}{x} = a e^{a \log x} e^{-\log x} = a e^{(a-1) \log x} = a x^{a-1}.$$

En particular, la derivada de $\sqrt[n]{x} = x^{1/n}$ (para $x > 0$) es

$$\frac{1}{n} x^{1/n-1} = \frac{1}{n} x^{-(n-1)/n} = \frac{1}{n \sqrt[n]{x^{n-1}}}.$$

Las derivadas de $\sinh x$ y $\cosh x$ se siguen fácilmente de que

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2}.$$

La regla del cociente nos da la derivada de la tangente hiperbólica. ■

Ejemplo Las reglas de derivación nos aseguran que cualquier función construida a partir de funciones elementales mediante sumas, productos, cocientes o composiciones es derivable en su dominio, salvo quizá donde una raíz se anula, pues las raíces están definidas en $x = 0$, pero sólo hemos probado que son derivables para $x > 0$. Además, dichas reglas nos permiten calcular fácilmente la derivada sin necesidad de calcular explícitamente ningún límite. Por ejemplo, la función

$$f(x) = \operatorname{sen}^5 \sqrt{x^2 + 3x}$$

es derivable en todos los puntos que cumplen $x^2 + 3x > 0$, y su derivada se puede calcular aplicando tres veces la regla de la cadena. En la primera tomamos $z(y) = y^5$ e $y(x) = \operatorname{sen} \sqrt{x^2 + 3x}$, con lo que $f(x) = z(y)$ y

$$f'(x) = 5y^4(x)y'(x) = 5 \operatorname{sen}^4 \sqrt{x^2 + 3x} y'(x).$$

En la segunda tomamos como $z(y) = \operatorname{sen} y$ e $y(x) = \sqrt{x^2 + 3x}$, con lo que la función compuesta $z(x)$ es la función $y(x)$ del paso anterior, luego

$$\begin{aligned} f'(x) &= 5y^4(x)y'(x) = 5 \operatorname{sen}^4 \sqrt{x^2 + 3x} \cos y(x) y'(x) \\ &= 5 \operatorname{sen}^4 \sqrt{x^2 + 3x} \cos \sqrt{x^2 + 3x} y'(x). \end{aligned}$$

En la tercera aplicación tomamos $z(y) = \sqrt{y}$ e $y(x) = x^2 + 3x$, con lo que llegamos al resultado final:

$$f'(x) = 5 \sin^4 \sqrt{x^2 + 3x} \cos \sqrt{x^2 + 3x} \frac{1}{2\sqrt{x^2 + 3x}} (2x + 3).$$

Notemos que, en la práctica, el cálculo de la derivada de una función compuesta se reduce a ir aplicando sucesivamente las reglas de derivación empezando por la función más externa y dejando invariantes las funciones interiores. Por ejemplo, derivar el exponente 5 inicial se reduce a poner el 5 multiplicando y sustituirlo por un 4 en el exponente, sin modificar nada más. ■

Derivación de una potencia Hemos visto como derivar la suma y el producto de dos funciones. Para potencias tenemos el teorema siguiente:

Teorema 1.7 Si $f :]a, b[\rightarrow]0, +\infty[$, $g :]a, b[\rightarrow \mathbb{R}$ son funciones derivables en un punto x , entonces f^g también es derivable en x y

$$(f^g)'(x) = f(x)^{g(x)} \left(g'(x) \log f(x) + g(x) \frac{f'(x)}{f(x)} \right).$$

DEMOSTRACIÓN: Basta tener en cuenta que, por definición

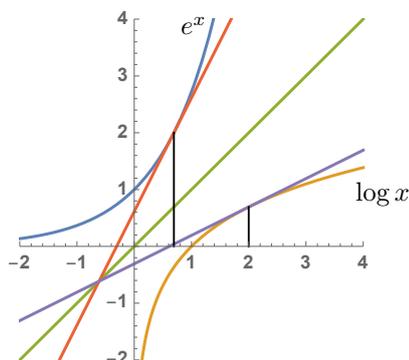
$$f(x)^{g(x)} = e^{g(x) \log f(x)},$$

luego, aplicando la regla de la cadena y la regla de derivación de un producto, concluimos que f^g también es derivable en x , así como que

$$(f^g)'(x) = e^{g(x) \log f(x)} \left(g'(x) \log f(x) + g(x) \frac{f'(x)}{f(x)} \right). \quad \blacksquare$$

Por ejemplo, si $f(x) = x^x$, entonces $f'(x) = x^x(1 + \log x)$.

El teorema de la función inversa La figura muestra las gráficas de las funciones e^x y $\log x$. El hecho de que sean mutuamente inversas se traduce en que las dos gráficas son simétricas respecto de la recta $y = x$. Por ejemplo, el punto $(2, \log 2)$ está en la gráfica del logaritmo, y su simétrico es el punto $(\log 2, 2) = (\log 2, e^{\log 2})$, que está en la gráfica de e^x :



Consecuentemente, las rectas tangentes en estos dos puntos son también simétricas respecto a la recta $y = x$, lo que se traduce en que sus pendientes son mutuamente inversas. Así, si la pendiente de la tangente de $\log x$ en $x = 2$ es $1/2$, la pendiente de la tangente en $\log 2$ de e^x tiene que ser 2 (y en efecto es así, pues la pendiente es $e^{\log 2} = 2$).

Esto es cierto en general siempre que tenemos dos funciones mutuamente inversas, digamos $y(x)$ y $x(y)$. Si la primera es derivable en x_0 , es decir, si su gráfica tiene tangente en el punto $(x_0, y(x_0))$, su inversa tiene que tener tangente en el punto $(y(x_0), x_0) = (y_0, x(y_0))$, y las pendientes tienen que ser mutuamente inversas:

$$\left. \frac{dx}{dy} \right|_{y_0} = \left(\left. \frac{dy}{dx} \right|_{x_0} \right)^{-1}.$$

La única excepción se da si una de las funciones tiene derivada nula en un punto, es decir, si su tangente es horizontal, en cuyo caso la inversa tendrá tangente vertical en el punto simétrico y no será derivable. Esto es lo que afirma el teorema de la función inversa:

Teorema 1.8 (Teorema de la función inversa) *Sea $f : A \rightarrow B$ una función derivable entre dos abiertos en \mathbb{K} con inversa continua $g : B \rightarrow A$. Si f es derivable en un punto x_0 y $f'(x_0) \neq 0$, entonces g es derivable en $y_0 = f(x_0)$ y $g'(y_0) = 1/f'(x_0)$.*

DEMOSTRACIÓN: Sea $h \neq 0$ tal que $y_0 + h \in B$. Sea $h' = g(y_0 + h) - g(y_0)$, con lo que $y_0 + h = f(x_0 + h')$. Entonces

$$\frac{g(y_0 + h) - g(y_0)}{h} = \frac{h' + x_0 - x_0}{y_0 + h - y_0} = \frac{h'}{f(x_0 + h') - f(x_0)} = \frac{1}{\frac{f(x_0 + h') - f(x_0)}{h'}}.$$

Sea $\{h_n\}_{n=0}^\infty$ una sucesión de términos no nulos que converja a 0. Entonces, como g es continua, tenemos que $h'_n = g(y_0 + h_n) - g(y_0)$ converge a 0, luego por la derivabilidad de f en x_0 tenemos que

$$\lim_n \frac{f(x_0 + h'_n) - f(x_0)}{h'_n} = f'(x_0),$$

luego

$$\lim_n \frac{g(y_0 + h_n) - g(y_0)}{h_n} = \frac{1}{f'(x_0)}$$

y esto prueba que existe

$$g'(y_0) = \lim_{h \rightarrow 0} \frac{g(y_0 + h) - g(y_0)}{h} = \frac{1}{f'(x_0)}. \quad \blacksquare$$

Ejemplos Con este teorema podemos calcular las derivadas de las funciones trigonométricas inversas. Por ejemplo, la función $\text{sen} :]-\pi, \pi[\rightarrow]-1, 1[$ es biyectiva y estrictamente creciente, por lo que su inversa

$$\text{arcsen} :]-1, 1[\rightarrow]-\pi, \pi[$$

es derivable en su dominio y, si $y = \arcsen x$, luego $x = \sen y$, tenemos que

$$\arcsen' x = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sen^2 y}} = \frac{1}{\sqrt{1 - x^2}}.$$

Similarmente se calculan las derivadas de las cinco últimas funciones de la tabla 1.1.

Derivación de funciones elementales complejas En realidad, el argumento que hemos dado para el cálculo de la derivada de la función e^x prueba sin cambio alguno que la exponencial compleja e^z es derivable en todo el plano complejo \mathbb{C} y que su derivada es ella misma. A su vez, las relaciones

$$\sen z = \frac{e^{iz} - e^{-iz}}{2i}, \quad \cos z = \frac{e^{iz} + e^{-iz}}{2}$$

implican inmediatamente que $\sen z$ y $\cos z$ son derivables en \mathbb{C} y que sus derivadas son $\cos z$ y $-\sen z$, respectivamente, de donde se sigue a su vez que la derivada de $\tan z$ es $1/\cos^2 z$, e igualmente concluimos que las funciones hiperbólicas son derivables en \mathbb{C} con las mismas derivadas que indica la tabla 1.1.

Similarmente, fijado cualquier logaritmo complejo de un número $a \in \mathbb{C}$ no nulo, podemos definir la exponencial $a^z = e^{z \log a}$, que resulta ser derivable en \mathbb{C} con la misma derivada que se indica en la tabla 1.1.

Si α es cualquier número real no nulo, podemos considerar el abierto H_α formado por todos los números complejos que no tienen argumento α . Es ciertamente un abierto, porque consta de todo el plano complejo menos una semirrecta cerrada de origen 0. También es abierto el conjunto A_α formado por los números complejos tales que $\alpha < \text{Im } z < \alpha + 2\pi$, y la función exponencial se restringe claramente a una biyección $A_\alpha \rightarrow H_\alpha$. Como su derivada es ella misma y no se anula, el teorema de la función inversa nos da que su inversa

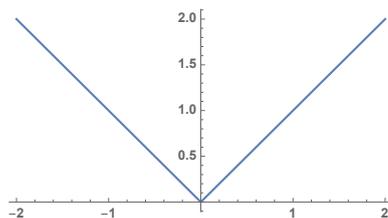
$$\log_\alpha : H_\alpha \rightarrow A_\alpha$$

es derivable y que su derivada es $\log'_\alpha z = 1/z$. En particular, si $\alpha = -\pi$ la función $\log_{-\pi}$ es la rama principal del logaritmo que definimos en [ITAn 6.3], y ahora sabemos que no sólo es continua, sino que es derivable.

En cualquier abierto $A \subset \mathbb{C}$ donde haya definido un logaritmo derivable, para cada $a \in \mathbb{C}$ podemos definir una exponenciación en A mediante $z^a = e^{a \log z}$, que resulta ser derivable con derivada az^{a-1} . ■

Ejemplos de funciones no derivables Hemos probado que toda función derivable en un punto es continua en dicho punto, luego cualquier función discontinua es un ejemplo de función no derivable. La idea subyacente es que si ampliamos una función alrededor de un punto en el que es discontinua nunca dejaremos de ver la discontinuidad, luego lo que veremos no se parecerá a una recta.

Entre las funciones continuas, el ejemplo más simple de función no derivable lo proporciona el valor absoluto $f(x) = |x|$:



Por más que ampliemos la gráfica de esta función alrededor del punto $(0, 0)$, nunca veremos algo parecido a una recta, sino que siempre veremos un “pico”, y eso implica que no es derivable. Explícitamente, si tratamos de calcular el límite

$$\lim_{h \rightarrow 0} \frac{|h| - |0|}{h} = \lim_{h \rightarrow 0} \frac{|h|}{h},$$

vemos que estamos tratando de calcular el límite de la función que vale -1 si $h < 0$ y 1 si $h > 0$, que claramente no tiene límite en 0 , luego la derivada no existe.

Pero también hay funciones continuas cuya gráfica no tiene “picos” y que no son derivables. Se trata de las funciones cuya recta tangente es vertical. Un ejemplo sencillo lo proporcionan las funciones $f(x) = \sqrt[n]{x}$ con n impar.

Notemos que la regla de derivación

$$f'(x) = \frac{1}{n \sqrt[n]{x^{n-1}}},$$

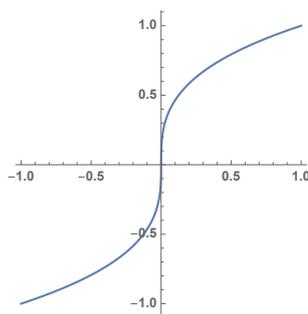
que en principio hemos demostrado para $x > 0$, también vale para $x < 0$, pues la restricción de f al intervalo $]-\infty, 0[$ puede calcularse como $f(x) = -f(-x)$, luego aplicando la regla de la cadena resulta que

$$f'(x) = -f'(-x)(-1) = f'(-x) = \frac{1}{n \sqrt[n]{(-x)^{n-1}}} = \frac{1}{n \sqrt[n]{x^{n-1}}},$$

pues $n - 1$ es par. En cambio, la función no es derivable en 0 , ya que la derivada debería ser

$$\lim_{h \rightarrow 0} \frac{\sqrt[n]{h} - 0}{h} = \lim_{h \rightarrow 0} \frac{1}{h^{1-1/n}} = \infty.$$

La figura siguiente muestra la gráfica de la función $f(x) = \sqrt[3]{x}$:



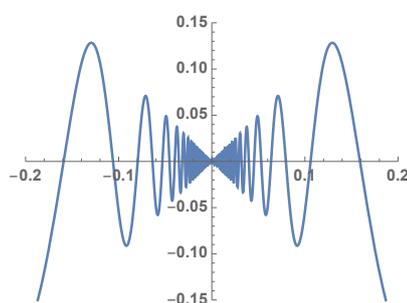
En ella vemos que la causa de la no derivabilidad es que la recta tangente en $x = 0$ es vertical. No obstante, podemos calcular dicha tangente si parametrizamos la gráfica, por ejemplo en la forma

$$r(x) = (x^n, x),$$

de modo que $r'(x) = (nx^{n-1}, 1)$ y, en particular, $r'(0) = (0, 1)$, que indica precisamente que la tangente en $(0, 0)$ es la recta (vertical) con vector director $(0, 1)$.

Un caso más “patológico” lo presenta la función

$$f(x) = \begin{cases} x \operatorname{sen} \frac{1}{x} & \text{si } x \neq 0, \\ 0 & \text{si } x = 0. \end{cases}$$



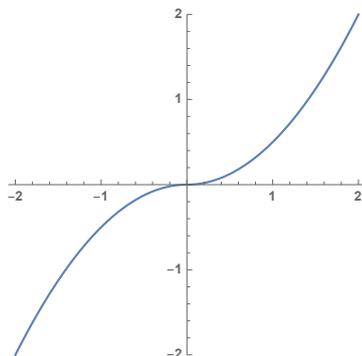
que es continua en $x = 0$, pero no es derivable porque, por mucho que ampliamos su gráfica alrededor de $(0, 0)$ nunca veremos nada parecido a una recta, sino que siempre veremos una curva oscilante. Analíticamente, la derivada debería ser

$$\lim_{h \rightarrow 0} \frac{h \operatorname{sen}(1/h)}{h} = \lim_{h \rightarrow 0} \operatorname{sen} \frac{1}{h},$$

pero es claro que este límite no existe.

Consideremos, por último, la función

$$f(x) = \begin{cases} x^2/2 & \text{si } x \geq 0, \\ -x^2/2 & \text{si } x \leq 0. \end{cases}$$



Usando la definición de derivada se prueba fácilmente que existe $f'(0) = 0$, con lo que, en general, f es derivable en \mathbb{R} y $f'(x) = |x|$. Así pues, se trata de un ejemplo de función derivable en \mathbb{R} cuya derivada no es derivable en todo \mathbb{R} . ■

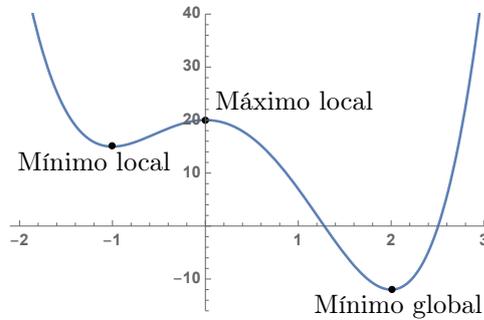
1.3 Máximos y mínimos

Las derivadas proporcionan un método para encontrar los puntos donde una función dada toma su valor máximo o mínimo. En realidad tenemos que hacer una precisión:

Definición 1.9 Un punto c es un *máximo local* (resp. *mínimo local*) de una función f si existe un $\delta > 0$ tal que $]c - \delta, c + \delta[$ está contenido su dominio y, para todo x en este intervalo, se cumple que $f(c) \geq f(x)$ (resp. $f(c) \leq f(x)$). Si se cumple $f(c) < f(x)$ (resp. $f(c) > f(x)$) para todo x en el intervalo se dice que el máximo (resp. mínimo) local es *estricto*.

Un punto donde una función tome el mayor (resp. el menor) valor posible es lo que se denomina un *máximo global* (resp. *mínimo global*). Obviamente todo máximo o mínimo global es también local, pero el recíproco no es cierto.

Por ejemplo, la figura muestra la función $f(x) = 3x^4 - 4x^3 - 12x^2 + 20$, que alcanza un mínimo global y no tiene máximo global (sino que tiende a $+\infty$ en $\pm\infty$, luego no hay ningún valor máximo que no supere). No obstante, tiene un máximo local y un segundo mínimo local distinto de su mínimo global.



Las derivadas nos permiten reconocer fácilmente los máximos y mínimos locales de una función:

Teorema 1.10 (Teorema de Fermat) Si una función f tiene un máximo o un mínimo local en un punto c y f es derivable en c , entonces $f'(c) = 0$.

DEMOSTRACIÓN: Supongamos que $f'(c) > 0$. Esto significa que

$$f'(c) = \lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} > 0,$$

luego por la definición de límite aplicada a $\epsilon = f'(c)$, existe un $\delta > 0$ (que podemos tomar menor que el δ dado por la definición de máximo o mínimo local) tal que si $0 < |h| < \delta$ se cumple que

$$\frac{f(c+h) - f(c)}{h} > 0.$$

Pero esto implica que si tomamos $h > 0$ entonces $f(c+h) > f(c)$, mientras que si $h < 0$ entonces $f(c+h) < f(c)$, en contradicción con que f tiene un máximo o un mínimo local en c . ■

Ejemplo En el caso de la función $f(x) = 3x^4 - 4x^3 - 12x^2 + 20$, al resolver la ecuación

$$f'(x) = 12x^3 - 12x^2 - 24x = 0$$

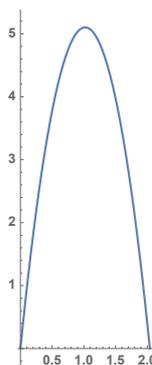
obtenemos las soluciones $x = -1, 0, 2$, que corresponden al máximo local ($x = 0$) y los dos mínimos locales ($x = -1, 2$) que muestra la gráfica anterior. ■

Nota Es importante tener presente que una función puede cumplir $f'(c) = 0$ sin necesidad de tener un máximo o un mínimo local en el punto c . Es el caso, por ejemplo, de la función $f(x) = x^3$. ■

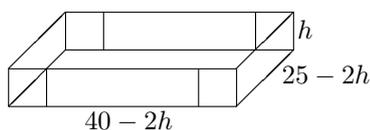
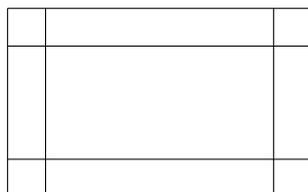
Ejemplo Más adelante veremos (página 30) que si lanzamos una piedra hacia arriba a una velocidad de v m/s, la altura de la piedra al cabo de t segundos será $h(t) = vt - 4.9t^2$ m. ¿Cuál será la máxima altura que alcanzará la piedra?

Dicha altura máxima será en particular un máximo local de la función $h(t)$ y, como ésta es derivable en todo \mathbb{R} , el teorema de Fermat nos asegura que el instante t en que se alcance la altura máxima debe cumplir $h'(t) = v - 9.8t = 0$, luego $t = v/9.8$ m y la altura máxima es $h(v/9.8) = 0.051v^2$.

La figura muestra el caso en el que la velocidad inicial es $v = 10$ m/s, por lo que la altura máxima alcanzada es de 5.1 m. Vemos que en los instantes en los que la piedra está subiendo la tangente a la gráfica tiene pendiente positiva, mientras que en los instantes en los que baja tiene pendiente negativa. En cambio, en el punto en el que la altura es máxima la pendiente es nula, y eso es precisamente lo que afirma el teorema de Fermat. ■



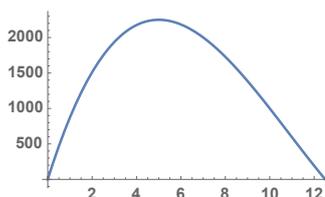
Ejemplo Tenemos una lámina de cartón de 40×25 cm y queremos usarla para construir una caja convirtiendo en solapas cuatro cuadrados de lado h como muestra la figura:



Determinar cuánto tiene que medir h para que el volumen de la caja sea máximo posible.

SOLUCIÓN: El volumen de la caja será

$$V = (40 - 2h)(25 - 2h)h = 1000h - 130h^2 + 4h^3$$

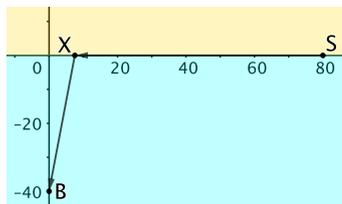


Por el teorema de Fermat, el valor de h que hace máximo el volumen debe cumplir que

$$V'(h) = 1000 - 260h + 12h^2 = 0.$$

Esta ecuación tiene dos soluciones, $h = 5$ y $h = 50/3 > 16$, pero la segunda es imposible, así que el máximo volumen se alcanza para $h = 5$, y es igual a $V(5) = 2250 \text{ cm}^3$. ■

Ejemplo *Un socorrista se encuentra en la orilla del mar en un punto S y ve que un bañista se está ahogando en un punto B situado a 80 m caminando por la costa y a 40 m de profundidad. El socorrista corre por la orilla hasta cierto punto x y luego se echa al agua. Si corre a 5 m/s y nada a 1 m/s, ¿en qué punto conviene que se eche al agua para llegar al bañista lo antes posible?*



SOLUCIÓN: En general, si el bañista se encuentra en $B = (0, y_b)$ y el socorrista en $S = (x_s, 0)$, el tiempo que tardará en llegar a al punto $X = (0, x)$ en que se echa al agua será $(x_s - x)/v_1$, donde v_1 es su velocidad en tierra, y luego tiene que nadar una distancia $\sqrt{x^2 + y_b^2}$, luego el tiempo total que tardará en llegar al bañista es

$$T(x) = \frac{x_s - x}{v_1} + \frac{\sqrt{x^2 + y_b^2}}{v_2},$$

donde v_2 es su velocidad nadando. Observemos que $T(80) = 89.44$ s, mientras que $T(0) = 56$ s, luego corriendo por la playa hasta situarse en el pie de la

perpendicular que pasa por el bañista le ahorra más de medio minuto. El punto óptimo para lanzarse al agua se obtiene igualando a 0 la derivada:

$$T'(x) = -\frac{1}{v_1} + \frac{x}{v_2\sqrt{x^2 + y_b^2}} = 0,$$

lo que equivale a $v_2\sqrt{x^2 + y_b^2} = v_1x$, luego $v_2^2(x^2 + y_b^2) = v_1^2x^2$, luego

$$x = \frac{k}{\sqrt{1 - k^2}} |y_b|,$$

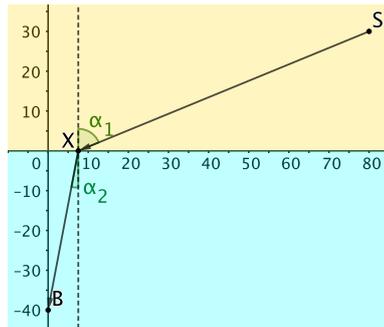
donde $k = v_2/v_1$. En nuestro caso,

$$x = \frac{0.2}{\sqrt{1 - 0.2^2}} 40 = 8.165 \text{ m}$$

y el tiempo correspondiente es $T(8.165) = 55.19$ s, luego optimizar su recorrido permite ahorrar al salvavidas casi 8 décimas de segundo. Ésta es la razón por la que los socorristas no necesitan saber derivar.

Observemos que la solución que hemos obtenido presupone que $k < 1$, es decir, que $v_2 < v_1$. En caso contrario es claro que la solución óptima sería lanzarse al agua inmediatamente. Por otro lado, vemos que x no depende de x_s , por lo que nos podría salir $x > x_s$, lo cual se debe a que en la definición de T deberíamos haber puesto $|x_s - x|$ en lugar de $x_s - x$, pues correr “hacia la derecha” no requiere un tiempo negativo. Si con otros datos saliera $x > x_s$, eso significaría que también conviene lanzarse al agua inmediatamente. ■

Ejemplo El ejemplo anterior es una versión simplificada de otro ejemplo que hemos planteado en la introducción (página xix). La diferencia consiste en suponer que el socorrista no está en la orilla del mar, sino, por ejemplo, 30 m tierra adentro:



En general, si el socorrista se encuentra en $S = (x_s, y_s)$, el tiempo que tarda en llegar hasta el bañista es

$$T(x) = \frac{\sqrt{(x_s - x)^2 + y_s^2}}{v_1} + \frac{\sqrt{x^2 + y_n^2}}{v_2}.$$

Tenemos que $T(0) = 236 \text{ s} = 3 \text{ min}, 56 \text{ s}$, pero el recorrido óptimo se obtiene igualando a 0 la derivada:

$$T'(x) = -\frac{x_s - x}{v_1 \sqrt{(x_s - x)^2 + y_s^2}} + \frac{x}{v_2 \sqrt{x^2 + y_n^2}},$$

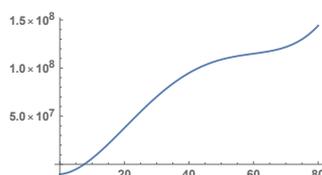
lo que nos da la ecuación

$$\frac{x_s - x}{\sqrt{(x_s - x)^2 + y_s^2}} = \frac{v_1}{v_2} \frac{x}{\sqrt{x^2 + y_n^2}}. \quad (1.1)$$

Con los datos del problema, tras elevar ambos miembros al cuadrado y simplificar, se reduce a

$$24x^4 - 3\,840x^3 + 174\,500x^2 + 256\,000x - 10\,240\,000 = 0.$$

No es fácil resolverla algebraicamente, pero podemos dibujar su gráfica, que muestra que tiene una única raíz real en el intervalo admisible para x :



La raíz puede ser aproximada numéricamente, y resulta ser $x = 7.521 \text{ m}$, que proporciona un tiempo $T(7.521) = 235.2 \text{ s}$, lo que supone, como antes, un ahorro de 8 décimas de segundo. ■

La ley de Snell En el ejemplo anterior no es fácil calcular x en general a partir de los datos del problema, pero la condición de minimalidad (1.1) tiene una interpretación geométrica muy simple, pues es equivalente a

$$\frac{\text{sen } \alpha_1}{\text{sen } \alpha_2} = \frac{v_1}{v_2}, \quad (1.2)$$

donde α_1 y α_2 , como muestra la figura, son los ángulos que forman con la perpendicular a la costa los segmentos que recorre el socorrista. Este hecho es interesante porque, como hemos visto en la introducción, demuestra que la ley de Snell sobre la refracción de la luz es consecuencia del principio de Fermat. ■

1.4 Propiedades de las funciones derivables

En esta sección demostraremos algunos teoremas sencillos sobre derivabilidad de funciones en los que se basa la potencia del cálculo diferencial. Aquí consideramos únicamente funciones de variable real. Al combinar el teorema de Fermat que hemos probado en la sección anterior con el teorema de Weierstrass [ITAn 3.31] obtenemos un hecho del que extraeremos varias consecuencias relevantes:

Teorema 1.11 (Teorema de Rolle) Sea $f : [a, b] \rightarrow \mathbb{R}$ una función continua en $[a, b]$ y derivable en $]a, b[$. Si $f(a) = f(b)$, entonces existe un $c \in]a, b[$ tal que $f'(c) = 0$.

DEMOSTRACIÓN: Por el teorema de Weierstrass, la función f alcanza un valor mínimo m y un valor máximo M . Si $m = M = f(a) = f(b)$, entonces f es constante y su derivada es nula, luego cualquier $c \in]a, b[$ cumple el teorema.

Supongamos que $m < M$. Entonces, o bien $m \neq f(a)$ o bien $M \neq f(a)$. Pongamos, por ejemplo, que $M \neq f(a)$. Sea $c \in [a, b]$ el punto donde $f(c) = M$. Como $M \neq f(a) = f(b)$, ha de ser $a < b < c$, y como f toma en c su valor máximo, en particular c es un máximo relativo de f , luego por el teorema de Fermat $f'(c) = 0$. ■

El teorema del valor medio Imaginemos que la función $f : [a, b] \rightarrow \mathbb{R}$ representa la posición en un instante t de un objeto puntual que se mueve sobre una recta. Durante el intervalo de tiempo $[a, b]$, ha pasado de la posición $f(a)$ a la posición $f(b)$, luego su desplazamiento global ha sido de $f(b) - f(a)$ (entendiendo que un signo positivo significa que ha terminado a la derecha del punto de partida y un signo negativo que ha terminado a la izquierda). No excluimos que el objeto haya podido avanzar y retroceder varias veces, de modo que la distancia recorrida puede haber sido en realidad mucho mayor, pero podemos decir que el balance final es que ha recorrido una distancia $f(b) - f(a)$ en un tiempo $b - a$, por lo que su *velocidad media* ha sido

$$v = \frac{f(b) - f(a)}{b - a}.$$

Esto hay que entenderlo como que si el objeto se hubiera movido en todo momento a esta velocidad habría terminado exactamente en el mismo punto en el que ha terminado. Pero en realidad puede haberse movido más rápidamente en algunas ocasiones y más lentamente en otras. Incluso podría haber estado parado parte del tiempo.

Pues bien, una de las propiedades más útiles de las funciones derivables es el llamado *teorema del valor medio*, que afirma que, sea cual sea el movimiento real que ha seguido el objeto, su velocidad en un determinado instante $a < c < b$ ha tenido que coincidir con la velocidad media. Explícitamente:

$$f(b) - f(a) = f'(c)(b - a).$$

En realidad vamos a demostrar algo más general: imaginemos que tenemos dos objetos moviéndose durante el mismo intervalo de tiempo, es decir, dos funciones $f, g : [a, b] \rightarrow \mathbb{R}$. Si el desplazamiento final de uno de ellos ha sido k veces mayor que el del otro, entonces en un determinado instante $a < c < b$ su velocidad ha tenido que ser exactamente k veces mayor que la del otro:

$$\frac{f'(c)}{g'(c)} = k = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Con precisión:

Teorema 1.12 (Teorema de Cauchy) Sean $f, g : [a, b] \rightarrow \mathbb{R}$ funciones continuas en $[a, b]$ y derivables en $]a, b[$. Entonces existe un $c \in]a, b[$ tal que

$$g'(c)(f(b) - f(a)) = f'(c)(g(b) - g(a)).$$

DEMOSTRACIÓN: Consideremos la función dada por

$$h(x) = f(x)(g(b) - g(a)) - g(x)(f(b) - f(a)).$$

Se cumple que $h(a) = h(b) = f(a)g(b) - g(a)f(b)$. Además h es continua en $[a, b]$ y derivable en $]a, b[$. Por el teorema de Rolle existe un punto $c \in]a, b[$ tal que $h'(c) = 0$, pero $h'(x) = f'(x)(g(b) - g(a)) - g'(x)(f(b) - f(a))$, luego

$$f'(c)(g(b) - g(a)) - g'(c)(f(b) - f(a)) = 0 \quad \blacksquare$$

Aplicando este teorema cuando $g(x) = x$ obtenemos:

Teorema 1.13 (Teorema del valor medio) Sea $f : [a, b] \rightarrow \mathbb{R}$ una función continua en $[a, b]$ y derivable en $]a, b[$. Entonces existe un $c \in]a, b[$ tal que

$$f(b) - f(a) = f'(c)(b - a).$$

Veamos algunas aplicaciones de este teorema:

Teorema 1.14 Si $f : [a, b] \rightarrow \mathbb{R}$ es continua en $[a, b]$, derivable en $]a, b[$ y su derivada es ≥ 0 , (resp. > 0 , ≤ 0 , < 0) en $]a, b[$, entonces f es creciente (resp. estrictamente creciente, decreciente, estrictamente decreciente) en $[a, b]$.

DEMOSTRACIÓN: Tomemos $a \leq x < y \leq b$. Por el teorema del valor medio, existe un $x < c < y$ de modo que

$$f(y) - f(x) = f'(c)(y - x),$$

y por la hipótesis sobre la derivada concluimos que $f(y) - f(x) \geq 0$ (resp. > 0 , ≤ 0 , < 0), lo que nos da la conclusión. \blacksquare

El lector puede ver una aplicación de este teorema en el análisis de los óvalos de Cassini en la sección B.6.

Sabemos que las funciones constantes tienen derivada nula. El teorema del valor medio nos da el recíproco:

Teorema 1.15 Si una función tiene derivada nula en todos los puntos de un intervalo abierto, entonces es constante.

DEMOSTRACIÓN: Sea f una función derivable en un intervalo A con derivada nula. Sean $a < b$ dos puntos cualesquiera de A . Entonces

$$f(b) - f(a) = f'(c)(b - a) = 0,$$

donde c es un punto de $]a, b[$. Por lo tanto f es constante. \blacksquare

Una consecuencia inmediata es el teorema siguiente, que afirma que una función derivable está unívocamente determinada por su derivada y su valor en un punto cualquiera.

Teorema 1.16 Si f y g son funciones derivables en un intervalo abierto y $f' = g'$, entonces existe un $k \in \mathbb{R}$ tal que $f = g + k$.

DEMOSTRACIÓN: La función $f - g$ tiene derivada nula, luego $f - g = k$. ■

Ejemplo: Movimiento rectilíneo uniforme Sea $e(t)$ la función que determina la posición en el instante t de un objeto que se mueve sobre una recta. Ya sabemos que su derivada $v(t) = e'(t)$ representa la velocidad a la que se mueve el objeto en cada instante. Si ésta resulta ser constante, $v(t) = v$, entonces una posible función $e(t)$ es $e^*(t) = vt$, pues ciertamente esta función tiene velocidad constante v . El teorema anterior nos asegura que $e(t) = vt + k$, para cierta constante k . Más precisamente, si sabemos que $e(t_0) = e_0$, sustituyendo estos valores y despejando k llegamos a que

$$e(t) = e_0 + v(t - t_0).$$

Ésta es la ecuación del *movimiento rectilíneo uniforme*, que permite calcular la posición de un objeto que se mueve en línea recta a velocidad constante conociendo su velocidad y su posición e_0 en un instante dado t_0 .

Por ejemplo, el 10 de enero de 1946 desde Belmar, un pueblo en el Estado de Nueva Jersey, se emitió una señal de radar que se reflejó en la Luna y fue captada de regreso unos 2.5 segundos más tarde. Teniendo en cuenta que la velocidad de la luz es de 299 792 km/s, el espacio recorrido por la señal de radar fue de 749 480 km, luego la Luna se encontraba a la mitad de dicha distancia, es decir, a 374 740 km de la Tierra. En realidad la distancia media es de 384 400 km, por lo que el tiempo transcurrido debió de ser, más precisamente, de unos 2.5644 s. ■

Ejemplo: Movimiento rectilíneo uniformemente acelerado En las condiciones del ejemplo precedente, supongamos que la velocidad del objeto no es constante. En tal caso podemos considerar su *aceleración*, que por definición es la derivada de la velocidad $a(t) = v'(t)$, que expresa la variación de la velocidad en cada instante t , exactamente en el mismo sentido en que la velocidad expresa la variación de la posición.

Consideremos el caso más simple en el que la aceleración es constante, es decir, $a(t) = a$. Entonces, exactamente el mismo razonamiento que hemos aplicado en el ejemplo anterior (cambiando la función $e(t)$ por la función $v(t)$) nos da que

$$v(t) = v_0 + a(t - t_0),$$

donde $v(t_0) = v_0$. Ahora bien, una función $e^*(t)$ cuya derivada es $v(t)$ es $e^*(t) = v_0t + \frac{1}{2}a(t - t_0)^2$, luego podemos concluir que $e(t) = k + v_0t + \frac{1}{2}a(t - t_0)^2$, para cierta constante k , y si sabemos además que $e(t_0) = e_0$, al sustituir y despejar k concluimos que

$$e(t) = e_0 + v_0t + \frac{1}{2}a(t - t_0)^2.$$

Estas dos ecuaciones determinan el *movimiento rectilíneo uniformemente acelerado*, y permiten calcular la posición y la velocidad en cada instante de un objeto que se mueve en línea recta con aceleración constante a conocidas su posición e_0 y su velocidad v_0 en un instante dado t_0 .

Por ejemplo, sucede que todo cuerpo que se mueve verticalmente cerca de la superficie terrestre experimenta una aceleración constante que puede variar ligeramente según el punto en que nos encontremos, pero que de media es⁶

$$g \approx 9.80665 \text{ m/s}^2.$$

Aquí hay que entender que si —como es habitual— fijamos un sistema de referencia en el que las posiciones más altas tienen coordenadas mayores, en las fórmulas g deberá incluirse con signo negativo:

$$e = e_0 + v_0(t - t_0) - \frac{1}{2}g(t - t_0)^2.$$

Esto explica las fórmulas que hemos empleado en los ejemplos de las páginas 6 y 23. Otro ejemplo:

Desde un puente sobre un río dejamos caer una piedra y oímos el chapoteo del agua al cabo de 3 segundos. ¿A qué altura está el puente sobre el nivel del agua?

Si consideramos que la piedra ha tardado 3 s en caer, la respuesta es que en ese tiempo ha recorrido

$$e(3) = -\frac{1}{2}g 3^2 = -44.1 \text{ m},$$

luego el puente está a 44.1 metros sobre la superficie del agua, pero si tenemos en cuenta que el sonido viaja a 340 m/s, un cálculo más exacto se obtiene llamando t_1 al tiempo que tarda la piedra en llegar al agua, de modo que en ese momento empieza a ascender una onda sonora desde la posición

$$h = -\frac{1}{2}g t_1^2,$$

que llegará a $h = 0$ en el instante $t = 3$, luego $0 = h + 340(3 - t_1)$. Por lo tanto:

$$340(3 - t_1) = -\frac{1}{2}g t_1^2$$

y la solución (positiva) de esta ecuación es $t_1 = 2.88$ s, y así la altura del puente resulta ser $-h = 40.65$ m. ■

⁶Notemos que si medimos las longitudes en metros (m) y el tiempo en segundos (s), la velocidad se mide en metros por segundo (m/s) y la aceleración en metros por segundo por segundo (m/s²).

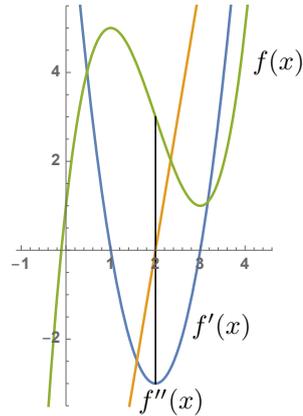
Funciones cóncavas y convexas Los resultados que hemos probado muestran ya que la derivada de una función nos proporciona información relevante sobre su gráfica. Por ejemplo, consideremos la función

$$f(x) = x^3 - 6x^2 + 9x + 1.$$

La figura muestra las gráficas de $f(x)$,

$$f'(x) = 3x^2 - 12x + 9 \quad \text{y} \quad f''(x) = 6x - 12.$$

En ella podemos comprobar lo que ya sabemos: que $f(x)$ crece en las zonas donde $f'(x)$ es positiva y decrece donde es negativa. La derivada se anula en $x = -1$ (donde f pasa de crecer a decrecer, por lo que tiene un máximo local) y en $x = 3$, donde pasa de decrecer a crecer, por lo que tiene un mínimo local. A su vez, $f'(x)$ pasa de decrecer a crecer en $x = 2$, que es donde se anula $f''(x)$, y en dicho punto toma su valor mínimo. Ahora bien el hecho de que la primera derivada pase de decrecer a crecer en el punto $x = 2$ tiene una repercusión visible en la gráfica de $f(x)$. En ese punto la función f pasa de ser cóncava a ser convexa, en el sentido que introducimos a continuación:



Definición 1.17 Diremos que una función $f :]a, b[\rightarrow \mathbb{R}$ es convexa (resp. cóncava) si para todo $a < x < y < b$ y todo $0 < \lambda < 1$ se cumple que

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y)$$

o

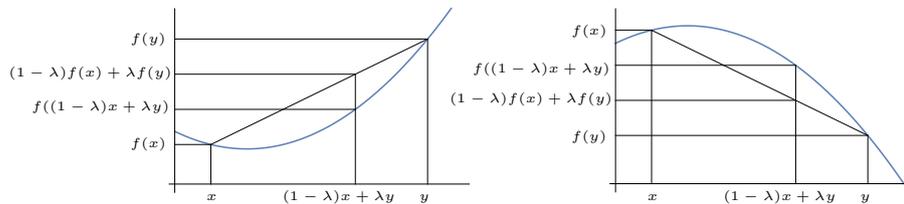
$$f((1 - \lambda)x + \lambda y) \geq (1 - \lambda)f(x) + \lambda f(y),$$

respectivamente.

Estas definiciones tienen una interpretación geométrica muy simple. Basta tener en cuenta que, cuando λ varía entre 0 y 1, la expresión $z = (1 - \lambda)x + \lambda y$ recorre todos los puntos intermedios entre x e y , e igualmente, $(1 - \lambda)f(x) + \lambda f(y)$ recorre todos los puntos intermedios entre $f(x)$ y $f(y)$. Más precisamente, la expresión

$$((1 - \lambda)x + \lambda y, (1 - \lambda)f(x) + \lambda f(y)) = (1 - \lambda)(x, f(x)) + \lambda(y, f(y))$$

recorre el segmento de extremos $(x, f(x))$ e $(y, f(y))$.



Así, la figura de la izquierda ilustra la definición de función convexa, que expresa que el segmento de extremos $(x, f(x))$ e $(y, f(y))$ está por encima de la gráfica de f , mientras que en el caso de una función cóncava (la figura de la derecha) el segmento está por debajo de la gráfica.

Es inmediato comprobar que una función f es cóncava si y sólo si $-f$ es convexa, y esto permite traducir inmediatamente a funciones cóncavas todos los resultados que vamos a demostrar para funciones convexas.

Si en $z = (1 - \lambda)x + \lambda y$ despejamos λ e incluimos la expresión en la definición de convexidad, obtenemos la versión equivalente: si $a < x < z < y < b$, entonces

$$f(z) \leq \frac{y-z}{y-x}f(x) + \frac{z-x}{y-x}f(y).$$

Operando:

$$yf(x) - zf(x) + zf(y) - xf(y) - yf(z) + xf(z) \geq 0.$$

Una comprobación rutinaria muestra ahora que esto equivale a cualquiera de las tres desigualdades siguientes:

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(y) - f(z)}{y - z}$$

(sólo hay que desarrollarlas y comprobar que llevan a la misma desigualdad precedente). Si llamamos

$$p(x, y) = \frac{f(y) - f(x)}{y - x},$$

podemos expresar así las desigualdades precedentes:

$$p(x, z) \leq p(x, y) \leq p(z, y),$$

es decir, que la pendiente del segmento que une dos puntos de la gráfica de f aumenta cuando se desplaza hacia la derecha cualquiera de ellos.

Veamos la relación entre la convexidad y las derivadas (el lector puede constatarla en el ejemplo dado al principio de este apartado):

Teorema 1.18 *Si $f :]a, b[\rightarrow \mathbb{R}$ es una función derivable y su derivada f' es creciente (decreciente), entonces f es convexa (cóncava).*

DEMOSTRACIÓN: Si $a < x < z < y < b$, por el teorema del valor medio existen $x < c < z < c' < y$ tales que

$$\frac{f(z) - f(x)}{z - x} = f'(c) \leq f'(c') = \frac{f(y) - f(z)}{y - z},$$

que es una de las tres caracterizaciones de la convexidad en términos de pendientes. ■

En particular, si f' es también derivable y su derivada f'' es ≥ 0 (resp. ≤ 0) en $]a, b[$, entonces f es convexa (resp. cóncava).

Otro hecho elemental es que una función convexa (cóncava) está siempre por encima (por debajo) de sus rectas tangentes:

Teorema 1.19 Si $f :]a, b[\rightarrow \mathbb{R}$ es una función convexa derivable en un punto x , entonces, para todo $y \in]a, b[$ se cumple que

$$f(x) + f'(x)(y - x) \leq f(y).$$

DEMOSTRACIÓN: Supongamos que $x < y$ y sea $x < z < y$. Entonces sabemos que

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x}.$$

Tomando el límite cuando $z \rightarrow x$ resulta que

$$f'(x) \leq \frac{f(y) - f(x)}{y - x}.$$

Si es $y < x$ usamos que

$$\frac{f(y) - f(x)}{y - x} = \frac{f(x) - f(y)}{x - y} \leq \frac{f(x) - f(z)}{x - z} = \frac{f(z) - f(x)}{z - x} \rightarrow f'(x)$$

y que $y - x < 0$, luego al despejarlo se invierte la desigualdad. ■

Una función cóncava o convexa no es necesariamente derivable (por ejemplo, la función $|x|$ es convexa), pero conviene observar que sí que es necesariamente continua:

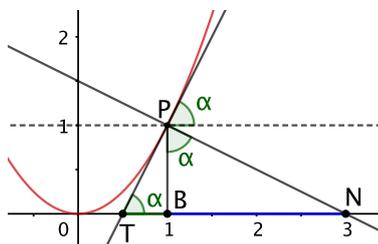
Teorema 1.20 Toda función cóncava o convexa $f :]a, b[\rightarrow \mathbb{R}$ es continua.

DEMOSTRACIÓN: Veamos que si $a < a' < b' < b$, para cada par de números $a' \leq x < y \leq b'$ existe un $C > 0$ tal que $|p(x, y)| \leq C$. En efecto, tomamos $a < a'' < a' < b' < b'' < b$ y observamos que

$$p(a', a'') \leq p(a', y) \leq p(x, y) \leq p(x, b') \leq p(b'', b'),$$

y basta tomar $C = \max\{|p(a', a'')|, |p(b'', b')|\}$, pues entonces $-C \leq p(x, y) \leq C$. Así pues, $|f(y) - f(x)| \leq C|y - x|$, y de aquí se sigue fácilmente la continuidad de f en $[a', b']$, luego en $]a, b[$. ■

Ejemplo: Subtangentes y subnormales Dada una función $y(x)$ y un punto $P = (x, y)$ en su gráfica, llamamos B al pie de la perpendicular al eje de abscisas desde P y T y N a los puntos donde la recta tangente y la recta normal (la perpendicular a la tangente) en P a la función cortan al eje.



El segmento \overline{BT} se llama *subtangente* y el segmento \overline{BN} se llama *subnormal* de la gráfica en el punto P .

Es claro que $y'(x) = \tan \alpha$, donde α es el ángulo indicado en la figura, luego

$$\overline{BT} = -\frac{y}{y'}, \quad \overline{BN} = yy', \quad (1.3)$$

donde el signo lo elegimos de modo que el resultado es positivo si el punto está a la derecha de B y negativo en caso contrario. Notemos que si $y' = 0$ entonces la tangente es horizontal y no corta al eje de abscisas, y por eso no está definida la subtangente.

Recordemos [ITAn 5.6] que todo punto (x, y) del plano puede expresarse como

$$(x, y) = \rho(\cos \theta, \text{sen } \theta),$$

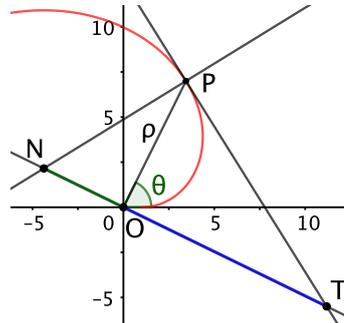
con $\rho \geq 0$ y $0 \leq \theta < 2\pi$. Los números ρ y θ se llaman *coordenadas polares* del punto (x, y) .

Podemos determinar una curva mediante una función $\rho(\theta)$. Concretamente, la formada por los puntos

$$(\rho(\theta) \cos \theta, \rho(\theta) \text{sen } \theta).$$

Se definen entonces la *subtangente polar* y la *subnormal polar* como los segmentos \overline{OT} y \overline{ON} determinados por el origen de coordenadas y los puntos T y N donde las rectas tangente y normal a la curva en P cortan a la perpendicular a OP que pasa por O .

Vamos a calcularlas en términos de la función $\rho(\theta)$.



Para ello consideramos el sistema de referencia que tiene como eje de abscisas la recta perpendicular a OP . Pongamos que P tiene coordenadas polares (ρ_0, θ_0) . En general, si un punto Q tiene coordenadas polares (ρ, θ) , respecto del nuevo sistema de referencia sus coordenadas polares son

$$(r, \phi) = (\rho, \theta - \theta_0 + \pi/2),$$

pues esta modificación de θ es la que hace que la coordenada polar θ_0 de P pase a ser $\phi_0 = \pi/2$. Así, en el nuevo sistema de referencia, la curva está determinada por la función $r(\phi) = \rho(\phi + \theta_0 - \pi/2)$. Notemos que, por la regla de la cadena,

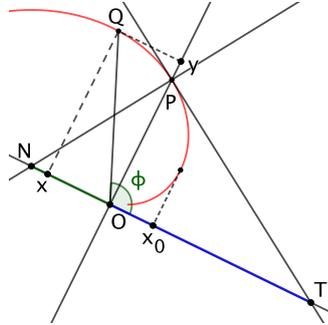
$$\frac{dr}{d\phi} = \frac{d\rho}{d\theta} \Big|_{\phi+\theta_0-\pi/2}.$$

Podemos suponer que la tangente a la curva en $\phi = \pi/2$ no es vertical, pues entonces la subtangente es nula y la subnormal no está definida. Esto equivale a que $x'(\pi/2) \neq 0$, luego, por continuidad, $x'(\phi)$ será siempre positiva o siempre negativa en puntos cercanos a $\pi/2$, luego $x(\phi)$ será creciente o decreciente alrededor de $\pi/2$ y podemos considerar la función inversa $\phi(x)$, de modo que

$$\frac{d\phi}{dx} = \left(\frac{dx}{d\phi}\right)^{-1}.$$

A su vez, podemos considerar la función compuesta $y(x) = y(\phi(x))$, definida para valores de x cercanos a 0 y cuya gráfica es al menos un segmento de la curva dada. Más aún, su subtangente y su subnormal (en el sentido cartesiano), son la subtangente y la subnormal polar de la curva dada.

Por ejemplo, en la figura, la función $y(x)$ sólo está definida para $x \leq x_0$:



Ahora podemos calcular la subtangente y la subnormal en P mediante las fórmulas cartesianas:

$$\begin{aligned} \overline{OT} &= -y(0) \left(\frac{dy}{dx}\Big|_0\right)^{-1} = -y(\pi/2) \left(\frac{dy}{d\phi}\Big|_{\pi/2} \frac{d\phi}{dx}\Big|_0\right)^{-1} \\ &= -y(\pi/2) \left(\frac{dy}{d\phi}\Big|_{\pi/2}\right)^{-1} \frac{dx}{d\phi}\Big|_{\pi/2}. \end{aligned}$$

Ahora derivamos en $(x, y) = (r \cos \phi, r \operatorname{sen} \phi)$:

$$\overline{OT} = -r(\pi/2) \operatorname{sen}(\pi/2) \frac{r'(\pi/2) \cos(\pi/2) - r(\pi/2) \operatorname{sen}(\pi/2)}{r'(\pi/2) \operatorname{sen}(\pi/2) + r(\pi/2) \cos(\pi/2)} = -\rho(\theta_0) \frac{-\rho(\theta_0)}{\rho'(\theta_0)}.$$

Con un cálculo análogo para la subnormal polar, concluimos que

$$\overline{OT} = \frac{\rho^2}{\rho'}, \quad \overline{ON} = y(0)y'(0) = -\rho'. \tag{1.4}$$

Así tenemos una interpretación geométrica de la derivada $d\rho/d\theta$ de una curva en coordenadas polares: es la longitud de la subnormal polar del punto P de coordenadas (ρ, θ) , entendiéndose que es positiva si la normal corta a la perpendicular a OP a la izquierda de O (entendiéndose que P está arriba de O) y negativa en caso contrario. ■

Derivación de series de funciones El criterio de mayoración de Weierstrass [ITAn 3.7] proporciona una condición suficiente para que una serie de funciones continuas converja a una función continua. Enunciamos ahora una variante que garantiza la derivabilidad de la suma, aunque por conveniencia pospondremos la prueba hasta el capítulo siguiente:

Teorema 1.21 Sea $f_n :]a, b[\rightarrow \mathbb{R}$ una sucesión de funciones derivables, de modo que exista un punto $x_0 \in]a, b[$ tal que la serie $\sum_{n=0}^{\infty} f_n(x_0)$ sea convergente. Sea $\{M_n\}_{n=0}^{\infty}$ una sucesión de números reales tal que $|f'_n(x)| \leq M_n$ para todo $x \in]a, b[$ y de modo que la serie $\sum_{n=0}^{\infty} M_n$ sea convergente. Entonces la serie $\sum_{n=0}^{\infty} f_n(x)$ converge en $]a, b[$ a una función derivable f tal que

$$f'(x) = \sum_{n=0}^{\infty} f'_n(x).$$

Admitiendo este resultado (que probaremos en la página 90) vamos a deducir de él una versión para series de potencias.

Según [ITAn 2.31] una serie de potencias

$$\sum_{n=0}^{\infty} a_n z^n,$$

donde los coeficientes a_n son números complejos, converge absolutamente en un disco abierto de centro 0 y radio $R \geq 0$, entendiéndose que puede ser $R = 0$ (en cuyo caso la serie sólo converge en $z = 0$) y también $R = +\infty$ (en cuyo caso la serie converge en todo el plano complejo).

Vamos a probar que la serie

$$\sum_{n=1}^{\infty} n a_n z^{n-1}$$

que resulta de derivar una serie dada término a término tiene exactamente el mismo radio de convergencia. Para ello empezamos observando que la serie

$$\sum_{n=0}^{\infty} n z^n$$

tiene radio de convergencia $R = 1$.

En efecto, es obvio que la serie diverge para $z = 1$, luego tiene que ser $R \leq 1$. Por otro lado, si $0 < r < 1$, tenemos que

$$\lim_n \frac{(n+1)r^{n+1}}{nr^n} = r < 1,$$

luego el criterio de D'Alembert [ITAn 2.12] implica que la serie converge en r , luego su radio de convergencia es exactamente $R = 1$.

Ahora llamemos R y R' a los radios de convergencia de las series

$$\sum_{n=1}^{\infty} a_n z^n \quad \text{y} \quad \sum_{n=1}^{\infty} n a_n z^{n-1},$$

respectivamente. Si $0 < r < R'$, tenemos que

$$|a_n R^n| = |a_n R^{n-1}| R \leq |n a_n z^{n-1}| r,$$

luego el criterio de comparación [ITAn 2.10] nos da la convergencia de

$$\sum_{n=1}^{\infty} |a_n r^n| \leq r \sum_{n=1}^{\infty} |n a_n r^{n-1}|,$$

luego $r \leq R$, y esto prueba que $R' \leq R$. Recíprocamente, si $r < R$, tomemos $r < s < R$. Hemos probado que la serie

$$\sum_{n=1}^{\infty} n \frac{r^n}{s^n}$$

es convergente, luego en particular nr^n/s^n tiende a 0, luego es una sucesión acotada. Sea $M > 0$ tal que $nr^n/s^n < M$ para todo n . Así $nr^n < Ms^n$, y el teorema de comparación nos da la convergencia de

$$\sum_{n=1}^{\infty} |n a_n r^{n-1}| = r^{-1} \sum_{n=1}^{\infty} |a_n| n r^n \leq (M/r) \sum_{n=1}^{\infty} |a_n| s^n.$$

Por lo tanto $r \leq R'$ y esto implica que $R = R'$.

Teorema 1.22 Si $f(x) = \sum_{n=0}^{\infty} a_n x^n$ es una serie de potencias con coeficientes reales y con radio de convergencia $R > 0$, entonces la serie $g(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$ tiene el mismo radio de convergencia, la función $f(x)$ es derivable en el intervalo $] -R, R[$ y en él se cumple que $f'(x) = g(x)$.

DEMOSTRACIÓN: Ya hemos probado que ambas series tienen el mismo radio de convergencia R y basta aplicar el teorema 1.21 a las funciones $f_n(x) = a_n x^n$. Basta probar que $f(x)$ es derivable en cualquier intervalo $] -r, r[$ con $0 < r < R$, y si $x \in] -r, r[$ tenemos que

$$|f_n(x)| = |a_n x^n| \leq |a_n r^n| = M_n$$

y

$$\sum_{n=0}^{\infty} M_n = \sum_{n=0}^{\infty} |a_n r^n|,$$

que es una serie convergente, porque $f(x)$ converge absolutamente en r . ■

Nota En 7.19 veremos que el teorema anterior es válido para series con coeficientes complejos y que la igualdad $f'(z) = g(z)$ es válida en realidad para todos los puntos del disco de convergencia común de las dos series. ■

En consecuencia, como todas las funciones definidas por series de potencias son derivables y sus derivadas también vienen definidas por series de potencias, concluimos que las funciones definidas por series de potencias son infinitamente derivables, es decir, que existen las derivadas primera, segunda, tercera, etc.

Ejemplo Si aplicamos el teorema anterior a la función exponencial:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

vemos que su derivada es

$$\sum_{n=1}^{\infty} \frac{nx^{n-1}}{n!} = \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x.$$

Tenemos así otra prueba de que e^x es derivable en \mathbb{R} y que su derivada es e^x . Similarmente, derivando las series de potencias correspondientes, podemos calcular las derivadas de las funciones $\sin x$ y $\cos x$. ■

Como aplicación del teorema anterior vamos a generalizar la fórmula del binomio de Newton. Para ello observamos que si en lugar de definir un número combinatorio como

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

lo definimos como

$$\binom{m}{n} = \frac{m(m-1)\cdots(m-n+1)}{n!} = \prod_{i=1}^n \frac{m+1-i}{i}$$

la definición es válida para todo número real \mathbb{R} , aunque no sea un número natural, y coincide con la usual cuando m es un número natural. Hay que entender que se cumple en general

$$\binom{m}{0} = 1.$$

Por ejemplo:

$$\binom{\sqrt{2}}{3} = \frac{\sqrt{2}(\sqrt{2}-1)(\sqrt{2}-2)}{1 \cdot 2 \cdot 3} = \frac{-8 + 5\sqrt{2}}{6} = -0.1548\dots$$

Teorema 1.23 Si α, x son números reales con $|x| < 1$, entonces

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n.$$

Notemos que, si α es un número natural, entonces $\binom{\alpha}{n} = 0$ cuando $n > \alpha$, por lo que la serie se reduce a la suma finita que proporciona el teorema del binomio de Newton.

DEMOSTRACIÓN: En primer lugar demostramos que la serie converge absolutamente usando el criterio de D'Alembert [ITAn 2.12]:

$$\begin{aligned} \lim_n \frac{|a_{n+1}|}{|a_n|} &= \lim_n \frac{|\alpha(\alpha-1)\cdots(\alpha-n)||x|^{n+1}}{(n+1)!} \cdot \frac{n!}{|\alpha(\alpha-1)\cdots(\alpha-n+1)||x|^n} \\ &= |x| \lim_n \frac{|\alpha-n|}{n+1} = |x| \lim_n \left| \frac{\alpha}{n+1} - \frac{n}{n+1} \right| = |x|. \end{aligned}$$

Por lo tanto, el radio de convergencia de la serie es $R = 1$. Consideramos la función $f :]-1, 1[\rightarrow \mathbb{R}$ dada por

$$f(x) = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n = \sum_{n=0}^{\infty} \frac{\prod_{k=0}^{n-1} (\alpha-k)}{n!} x^n.$$

Por el teorema anterior

$$f'(x) = \sum_{n=1}^{\infty} \frac{\prod_{k=0}^{n-1} (\alpha-k)}{(n-1)!} x^{n-1}.$$

Por lo tanto,

$$\begin{aligned} (1+x)f'(x) &= \sum_{n=1}^{\infty} \frac{\prod_{k=0}^{n-1} (\alpha-k)}{(n-1)!} x^{n-1} + \sum_{n=1}^{\infty} \frac{\prod_{k=0}^{n-1} (\alpha-k)}{(n-1)!} x^n = \\ &= \alpha + \sum_{n=1}^{\infty} \frac{\prod_{k=0}^n (\alpha-k)}{n!} x^n + \sum_{n=1}^{\infty} \frac{\prod_{k=0}^{n-1} (\alpha-k)}{(n-1)!} x^n = \\ &= \alpha + \sum_{n=1}^{\infty} \frac{\prod_{k=0}^n (\alpha-k)}{(n-1)!} \left(\frac{1}{n} + \frac{1}{\alpha-n} \right) x^n = \alpha + \sum_{n=1}^{\infty} \frac{\prod_{k=0}^n (\alpha-k)}{(n-1)!} \frac{\alpha}{n(\alpha-n)} x^n = \\ &= \alpha + \alpha \sum_{n=1}^{\infty} \frac{\prod_{k=0}^{n-1} (\alpha-k)}{n!} x^n = \alpha f(x). \end{aligned}$$

Así pues, $\alpha f(x) = (1+x)f'(x)$, y si llamamos $g(x) = f(x)(1+x)^{-\alpha}$, tenemos que

$$\begin{aligned} g'(x) &= f'(x)(1+x)^{-\alpha} + f(x)(-\alpha)(1+x)^{-\alpha-1} \\ &= f'(x)(1+x)^{-\alpha} - (1+x)f'(x)(1+x)^{-\alpha-1} = f'(x)(1+x)^{-\alpha} - f'(x)(1+x)^{-\alpha} = 0. \end{aligned}$$

El teorema 1.15 nos da ahora que $g(x)$ es constante, es decir, que existe un número real c tal que $f(x) = c(1+x)^\alpha$, pero haciendo $x = 0$ obtenemos que la constante es $c = 1$. ■

Por ejemplo:

$$\begin{aligned}\sqrt{1+x} &= 1 + \frac{1/2}{1}x + \frac{1/2(1/2-1)}{2}x^2 + \frac{1/2(1/2-1)(1/2-2)}{6}x^3 + \dots \\ &= +\frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \dots\end{aligned}$$

Podría pensarse que el teorema 1.21 tendría que poder probarse con la hipótesis $|f_n(x)| \leq M_n$ en lugar de $|f'_n(x)| \leq M_n$. Sin embargo, vamos a ver que dicha hipótesis garantiza ciertamente la continuidad de la serie (por el criterio de mayoración de Weierstrass) pero no su derivabilidad (compárese en cambio con el teorema 7.18):

Ejemplo (Weierstrass) Sea $0 < a < 1$ y $b > 1$ un entero impar de modo que $ab > 1 + 3\pi/2$. Entonces la función

$$W(x) = \sum_{k=0}^{\infty} a^k \cos(b^k \pi x)$$

es continua en \mathbb{R} , pero no es derivable en ningún punto.

Notemos en primer lugar que

$$|a^k \cos(b^k \pi x)| \leq a^k,$$

y la serie geométrica $\sum_{k=0}^{\infty} a^k$ es convergente, luego el criterio de mayoración de Weierstrass implica que la serie converge uniformemente en \mathbb{R} a una función continua.

En primer lugar observamos que, por el teorema del valor medio, para todo $x, y \in \mathbb{R}$ se cumple

$$|\cos x - \cos y| = |\sin \xi(x-y)| \leq |x-y|.$$

También es claro que si k es un entero, entonces

$$\cos(k\pi + x) = (-1)^k \cos x.$$

Fijemos $x_0 \in \mathbb{R}$ y sea m un número natural. Sea r_m el entero más próximo a $b^m x_0$, de modo que

$$b^m x_0 - \frac{1}{2} \leq r_m \leq b^m x_0 + \frac{1}{2}.$$

Sea

$$x_m = \frac{r_m + 1}{b^m}.$$

Así

$$x_0 = \frac{b^m x_0}{b^m} < \frac{b^m x_0 - \frac{1}{2} + 1}{b^m} \leq x_m,$$

$$x_k \leq \frac{b^k x_0 + \frac{1}{2} + 1}{b^k} = x_0 + \frac{3}{2b^k}.$$

Por consiguiente:

$$x_0 < x_m < x_0 + \frac{3}{2b^m},$$

de donde concluimos que $\lim_m x_m = x_0$. Consideramos ahora

$$W(x_m) - W(x_0) = \sum_{k=0}^{\infty} a^k (\cos(b^k \pi x_m) - \cos(b^k \pi x_0)) = A + B,$$

donde

$$\begin{aligned} A &= \sum_{k=0}^{m-1} a^k (\cos(b^k \pi x_m) - \cos(b^k \pi x_0)), \\ B &= \sum_{k=m}^{\infty} a^k (\cos(b^k \pi x_m) - \cos(b^k \pi x_0)). \end{aligned}$$

Por una parte tenemos que

$$\begin{aligned} |A| &\leq \sum_{k=0}^{m-1} a^k |\cos(b^k \pi x_m) - \cos(b^k \pi x_0)| \leq \sum_{k=0}^{m-1} a^k b^k \pi (x_m - x_0) \\ &\leq \pi (x_m - x_0) \frac{(ab)^m - 1}{ab - 1} \leq \frac{\pi (ab)^m}{ab - 1} (x_m - x_0). \end{aligned}$$

Por otra parte

$$\cos(b^k \pi x_m) = \cos(b^k \pi \frac{r_m + 1}{b^m}) = \cos(b^{k-m} (r_m + 1) \pi).$$

Para $k \geq m$ tenemos que $b^{k-m} (r_m + 1)$ es entero, luego, teniendo en cuenta que b es impar:

$$\cos(b^k \pi x_m) = (-1)^{b^{k-m} (r_m + 1)} = -(-1)^{r_m}.$$

Además

$$\cos(b^k \pi x_0) = \cos(b^k \pi \frac{r_m + b^m x_0 - r_m}{b^m}) = \cos(b^{k-m} r_m \pi + b^{k-m} z_m \pi),$$

donde $z_m = b^m x_0 - r_m$.

De nuevo, si $k \geq m$ tenemos que $b^{k-m} r_m$ es un entero, luego

$$\cos(b^k \pi x_0) = (-1)^{b^{k-m} r_m} \cos(b^{k-m} z_m \pi) = (-1)^{r_m} \cos(b^{k-m} z_m \pi).$$

Esto nos da la expresión siguiente para B :

$$\begin{aligned} B &= \sum_{k=m}^{\infty} a^k (-(-1)^{r_m} - (-1)^{r_m} \cos(b^{k-m} z_m \pi)) \\ &= -(-1)^{r_m} \sum_{k=m}^{\infty} a^k (1 + \cos(b^{k-m} z_m \pi)). \end{aligned}$$

Como todos los términos de la serie son no negativos, podemos concluir que

$$|B| \geq a^m(1 + \cos(z_m\pi)).$$

Más aún, tenemos que $|z_m| = |b^m x_0 - r_m| \leq 1/2$, luego $|z_m\pi| \leq \pi/2$, luego el coseno de la última expresión es no negativo, luego $|B| \geq a^m$.

Ahora usamos que $x_m - x_0 \leq 3/2b^m$, luego $2b^m(x_m - x_0)/3 \leq 1$, con lo que podemos concluir que

$$|B| \geq \frac{2a^m b^m}{3}(x_m - x_0).$$

Combinando las desigualdades que hemos obtenido resulta que

$$\begin{aligned} |A + B| &\geq |B| - |A| \geq \frac{2a^m b^m}{3}(x_m - x_0) - \frac{\pi(ab)^m}{ab - 1}(x_m - x_0) \\ &= (ab)^m \left(\frac{2}{3} - \frac{\pi}{ab - 1} \right) (x_m - x_0). \end{aligned}$$

Por consiguiente:

$$\left| \frac{W(x_m) - W(x_0)}{x_m - x_0} \right| \geq (ab)^m \left(\frac{2}{3} - \frac{\pi}{ab - 1} \right).$$

La hipótesis del teorema garantiza que $ab > 1$ y que el segundo factor es positivo, luego

$$\lim_m \left| \frac{W(x_m) - W(x_0)}{x_m - x_0} \right| = +\infty,$$

lo que implica que W no es derivable en x_0 . ■

Nota Observemos que si derivamos término a término la serie que define a W obtenemos la serie

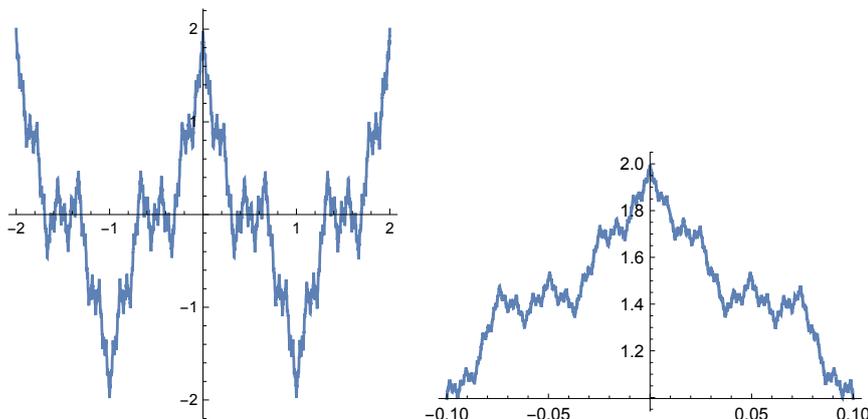
$$-\pi \sum_{k=0}^{\infty} (ab)^k \operatorname{sen}(b^k \pi x),$$

luego si $ab < 1$ podemos aplicar el teorema de mayoración de Weierstrass para probar que la serie converge uniformemente en \mathbb{R} , y entonces el teorema 1.21 implica que, de hecho, existe

$$W'(x) = -\pi \sum_{k=0}^{\infty} (ab)^k \operatorname{sen}(b^k \pi x).$$

Por lo tanto, para que la función de Weierstrass no sea derivable es necesario que $ab \geq 1$. Hardy demostró que esta condición es también suficiente. ■

La figura muestra la gráfica de W para $a = 0.5$ y $b = 3$ en el intervalo $[-2, 2]$, así como una ampliación alrededor de 0:



1.5 La regla de L'Hôpital

Demostremos ahora diversas formas del teorema conocido como “regla de l'Hôpital” si bien, como hemos explicado en la introducción fue descubierto por Johann Bernoulli:

Teorema 1.24 (Regla de L'Hôpital) Sean $f, g :]a, b[\rightarrow \mathbb{R}$ funciones derivables tales que $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = 0$ y de modo que g y g' no se anulen en $]a, b[$. Si existe

$$\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L,$$

entonces también existe

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L.$$

DEMOSTRACIÓN: Extendamos f y g al intervalo $[a, b[$ estableciendo que $f(a) = g(a) = 0$. Así siguen siendo continuas.

Si $a < x < b$, por el teorema de Cauchy existe un punto $c \in]a, x[$ tal que

$$(f(x) - f(a))g'(c) = (g(x) - g(a))f'(c),$$

o sea, $f(x)g'(c) = g(x)f'(c)$, y como $g(x) \neq 0 \neq g'(c)$, podemos escribir

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)}. \quad (1.5)$$

Por definición de límite, si $\epsilon > 0$, existe un $\delta > 0$ tal que si $0 < c - a < \delta$, entonces

$$\left| \frac{f'(c)}{g'(c)} - L \right| < \epsilon. \quad (1.6)$$

Así tenemos que si $0 < x - a < \delta$, existe un $c \in]a, x[$ que cumple (1.5) y (1.6). Por consiguiente, para todo $x \in]a, a + \delta[$ se cumple

$$\left| \frac{f(x)}{g(x)} - L \right| < \epsilon.$$

Esto significa que

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L. \quad \blacksquare$$

Obviamente la regla de L'Hôpital también es válida cuando en las hipótesis cambiamos a por b . Combinando las dos versiones obtenemos la regla de L'Hôpital para funciones definidas en intervalos $]a - \epsilon, a + \epsilon[\setminus \{a\}$ y tomando límites en a (si existe el límite del cociente de derivadas, existen los límites por la derecha y por la izquierda y coinciden, por los casos correspondientes de la regla, existen los límites de los cocientes de las funciones por ambos lados y coinciden, luego existe el límite y coincide con el de las derivadas).

Ejemplos Usando la regla de L'Hôpital podemos calcular, por ejemplo:

$$\lim_{x \rightarrow 1} \frac{\operatorname{sen} \pi x}{\ln x} = \lim_{x \rightarrow 1} \frac{\pi \cos \pi x}{1/x} = -\pi.$$

A veces hace falta aplicar la regla varias veces seguidas:

$$\lim_{x \rightarrow 0} \frac{\operatorname{sen} x - x}{x^3} = \lim_{x \rightarrow 0} \frac{\cos x - 1}{3x^2} = \lim_{x \rightarrow 0} \frac{\operatorname{sen} x}{6x} = \lim_{x \rightarrow 0} \frac{\cos x}{6} = \frac{1}{6}. \quad \blacksquare$$

Los teoremas siguientes demuestran otras variantes de la regla de L'Hôpital:

Teorema 1.25 (Regla de L'Hôpital) Sean $f, g :]a, +\infty[\rightarrow \mathbb{R}$ dos funciones derivables tales que $\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow +\infty} g(x) = 0$ y de modo que g y g' no se anulan en $]a, +\infty[$. Si existe

$$\lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)} = L,$$

entonces también existe

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = L.$$

DEMOSTRACIÓN: Consideremos $F(x) = f(1/x)$ y $G(x) = g(1/x)$, definidas en $]0, 1/a[$. Claramente F y G son continuas, y $\lim_{x \rightarrow 0} F(x) = \lim_{x \rightarrow 0} G(x) = 0$. Además por la regla de la cadena son funciones derivables y sus derivadas son

$$F'(x) = -\frac{f'(1/x)}{x^2}, \quad G'(x) = -\frac{g'(1/x)}{x^2}.$$

También es claro que ni G ni G' se anulan en su dominio y

$$\frac{F'(x)}{G'(x)} = \frac{f'(1/x)}{g'(1/x)},$$

luego existe

$$\lim_{x \rightarrow 0} \frac{F'(x)}{G'(x)} = L.$$

El caso ya probado de la regla de L'Hôpital nos da ahora que también existe

$$\lim_{x \rightarrow 0} \frac{F(x)}{G(x)} = L.$$

Obviamente entonces

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = L.$$

■

Igualmente se prueba la regla de L'Hôpital para funciones definidas en intervalos $]-\infty, b[$ y cuando x tiende a $-\infty$.

Así pues, si tenemos una indeterminación de tipo $0/0$ y al derivar numerador y denominador podemos calcular el límite, la función original tiene ese mismo límite. Ahora veremos que la regla de L'Hôpital es aplicable también a indeterminaciones del tipo ∞/∞ .

Teorema 1.26 (Regla de L'Hôpital) Sean $f, g :]a, +\infty[\rightarrow \mathbb{R}$ dos funciones derivables tales que $\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow +\infty} g(x) = \infty$ y de modo que g y g' no se anulan en $]a, +\infty[$. Si existe

$$\lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)} = L,$$

entonces también existe

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = L.$$

DEMOSTRACIÓN: Por definición de límite, dado $\epsilon > 0$, existe un $M > a$ tal que si $x > M$ entonces

$$\left| \frac{f'(x)}{g'(x)} - L \right| < \epsilon.$$

Por el teorema de Cauchy, si $x > M$, existe un $y \in]M, x[$ de modo que

$$\frac{f(x) - f(M)}{g(x) - g(M)} = \frac{f'(y)}{g'(y)},$$

luego

$$\left| \frac{f(x) - f(M)}{g(x) - g(M)} - L \right| < \epsilon.$$

(Notemos que, como g' no se anula, la función g es monótona, luego el denominador es no nulo).

Puesto que $\lim_{x \rightarrow +\infty} f(x) = \infty$, existe un $N > M$ tal que si $x > N$ entonces $|f(x)| > |f(M)|$, y en particular $f(x) - f(M) \neq 0$. Por ello, para $x > N$ podemos escribir

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(M)}{g(x) - g(M)} \frac{f(x)}{f(x) - f(M)} \frac{g(x) - g(M)}{g(x)}.$$

Si en los dos últimos factores dividimos numerador y denominador entre $f(x)$ y $g(x)$ respectivamente, queda claro que tienden a 1 cuando $x \rightarrow +\infty$, luego tomando N suficientemente grande podemos suponer que si $x > N$ entonces el producto de ambos dista de 1 menos de ϵ . De este modo, para x suficientemente grande, el cociente $f(x)/g(x)$ se puede expresar como producto de dos números reales, uno arbitrariamente próximo a L y otro arbitrariamente próximo a 1. De la continuidad del producto se sigue que

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = L. \quad \blacksquare$$

Naturalmente la regla de L'Hôpital también es válida en el caso ∞/∞ cuando $x \rightarrow -\infty$. El mismo argumento que nos ha permitido pasar del caso finito al caso infinito en la indeterminación $0/0$ nos permite pasar ahora al caso finito. Es fácil probar:

Teorema 1.27 (Regla de L'Hôpital) Sean $f, g :]a, b[\rightarrow \mathbb{R}$ derivables tales que $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = \infty$ y de modo que g y g' no se anulan en $]a, b[$. Si existe

$$\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L,$$

entonces también existe

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L.$$

También se cumple la versión correspondiente cuando x tiende a b y cuando x tiende a un punto por la izquierda y la derecha a la vez.

Contraejemplos El argumento siguiente es incorrecto: Como

$$\lim_{x \rightarrow +\infty} \frac{x + \operatorname{sen} x}{x} = \lim_{x \rightarrow +\infty} \frac{1 + \cos x}{1}$$

y el segundo límite no existe, el primero tampoco.

Esto es falso, pues

$$\lim_{x \rightarrow +\infty} \frac{x + \operatorname{sen} x}{x} = \lim_{x \rightarrow +\infty} 1 + \frac{\operatorname{sen} x}{x} = 2.$$

El error está en que la regla de L'Hôpital sólo afirma que si existe el límite del cociente de las derivadas, éste coincide con el límite del cociente inicial, pero no dice que si el límite del cociente de las derivadas no existe no pueda existir el límite del cociente inicial. El ejemplo se puede adaptar a casos análogos para las otras formas de la regla.

También es incorrecto el cálculo siguiente:

$$\begin{aligned} \lim_{x \rightarrow +\infty} \frac{x + \operatorname{sen} x \cos x}{(x + \operatorname{sen} x \cos x)e^{\operatorname{sen} x}} &= \\ \lim_{x \rightarrow +\infty} \frac{2 \cos^2 x}{2 \cos^2 x e^{\operatorname{sen} x} + (x + \operatorname{sen} x \cos c) \cos x e^{\operatorname{sen} x}} &= 0. \end{aligned}$$

Por el contrario:

$$\lim_{x \rightarrow +\infty} \frac{x + \operatorname{sen} x \cos x}{(x + \operatorname{sen} x \cos x)e^{\operatorname{sen} x}} = \lim_{x \rightarrow +\infty} e^{-\operatorname{sen} x}$$

y es claro que este límite no existe. El fallo está en que no se cumple la hipótesis de que g' no se anule en un intervalo $]a, +\infty[$, por lo que la regla de L'Hôpital no se puede aplicar. ■

Ejemplo Aplicando la regla de L'Hôpital vemos que

$$\lim_{x \rightarrow 0} x \log x = \lim_{x \rightarrow 0} \frac{\log x}{1/x} = \lim_{x \rightarrow 0} \frac{1/x}{-1/x^2} = \lim_{x \rightarrow 0} -x = 0.$$

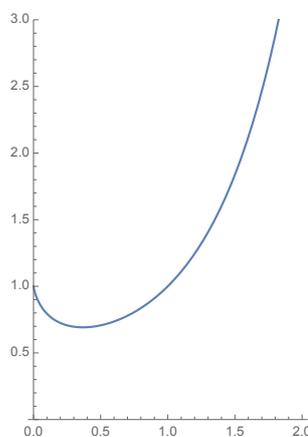
Por consiguiente,

$$\lim_{x \rightarrow 0} x^x = \lim_{x \rightarrow 0} e^{x \log x} = e^0 = 1.$$

Así pues, la función $f(x) = x^x$ es continua en el intervalo $[0, +\infty[$ si entendemos que en 0 toma el valor $0^0 = 1$. Su derivada es

$$f'(x) = x^x(\log x + 1),$$

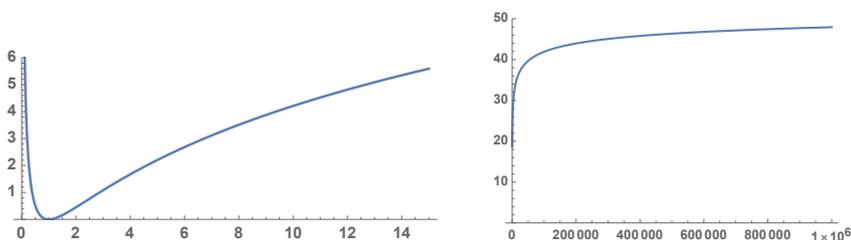
y tiende a $-\infty$ en $a = 0$. Además se anula en $a = 1/e$, luego ése es el punto donde alcanza su valor mínimo. ■



Ejemplo La regla de L'Hôpital nos da también que, si $\alpha > 0$, entonces

$$\begin{aligned} \lim_{x \rightarrow +\infty} \frac{\log^2 x}{x^\alpha} &= \lim_{x \rightarrow +\infty} \frac{2 \log x \cdot 1/x}{\alpha x^{\alpha-1}} = \\ \lim_{x \rightarrow +\infty} \frac{2 \log x}{\alpha x^\alpha} &= \lim_{x \rightarrow +\infty} \frac{2/x}{\alpha^2 x^{\alpha-1}} = \lim_{x \rightarrow +\infty} \frac{2}{\alpha^2 x^\alpha} = 0. \end{aligned}$$

Sin embargo, nadie diría que esto es así viendo la gráfica de esta función, por ejemplo para $\alpha = 0.1$:



La derivada de $f(x) = \log^2 x/x^\alpha$ es

$$f'(x) = \frac{2x^{\alpha-1} \log x - \alpha x^{\alpha-1} \log^2 x}{x^{2\alpha}} = \frac{x^{\alpha-1} \log x (2 - \alpha \log x)}{x^{2\alpha}},$$

que se anula cuando $\log x = 0$ (o sea, en $x = 1$) y cuando $2 - \alpha \log x = 0$, lo cual sucede en $x^* = e^{2/\alpha}$.

Claramente, en $x = 1$ la función toma su valor mínimo, que es $f(1) = 0$, mientras que en x^* la derivada pasa de ser positiva a ser negativa, luego el punto es un máximo local, y el valor de la función es

$$f(e^{2/\alpha}) = \frac{(2/\alpha)^2}{e^2} = \left(\frac{2}{\alpha e}\right)^2.$$

Por ejemplo, para $\alpha = 0.1$, el valor máximo se alcanza en

$$x^* = e^{20} = 485\,165\,195.4, \quad \text{donde } f(x^*) = 54.13,$$

y sólo a partir de ahí empieza a decrecer hacia 0, pero si queremos que descienda hasta un punto donde, por ejemplo, $f(x) = 0.01$, necesitamos llegar al menos hasta $x = 3.97 \cdot 10^{39}$. Esto hace que algunos resultados que involucran límites describen hechos que quedan mucho más allá de nuestra capacidad de cálculo. Veamos un ejemplo:

Teorema 1.28 *Para cada número natural n , sea $d(n)$ el número de divisores de n . Entonces, para todo $\epsilon > 0$ existe un n_0 tal que si $n \geq n_0$, entonces*

$$\log d(n) \leq (1 + \epsilon) \log 2 \frac{\log n}{\log \log n}.$$

DEMOSTRACIÓN: Sea $n = p_1^{e_1} \cdots p_r^{e_r}$ la descomposición en factores primos de un número natural n y consideremos un $\delta > 0$ que luego especificaremos. Cada divisor de n es de la forma $p_1^{f_1} \cdots p_r^{f_r}$ con $0 \leq f_i \leq e_i$, por lo que, claramente,

$$d(n) = (e_1 + 1) \cdots (e_r + 1).$$

Entonces

$$\frac{d(n)}{n^\delta} = \prod_{i=1}^r \frac{e_i + 1}{p_i^{e_i \delta}}.$$

Usando la desigualdad $1 + x \leq e^x$ vemos que

$$e_i \delta \log 2 \leq e^{e_i \delta \log 2} = 2^{e_i \delta} \leq p_i^{e_i \delta}.$$

$$\frac{e_i + 1}{p_i^{e_i \delta}} \leq \frac{e_i + p_i^{e_i \delta}}{p_i^{e_i \delta}} = 1 + \frac{e_i}{p_i^{e_i \delta}} \leq 1 + \frac{1}{\delta \log 2} \leq e^{1/(\delta \log 2)}.$$

Por otra parte, si $p_i \geq 2^{1/\delta}$, entonces $p_i^\delta \geq 2$ y

$$\frac{e_i + 1}{p_i^{e_i \delta}} \leq \frac{e_i + 1}{2^{e_i}}.$$

Por lo tanto,

$$\frac{d(n)}{n^\delta} = \prod_{p_i < 2^{1/\delta}} \frac{e_i + 1}{p_i^{e_i \delta}} \prod_{p_i \geq 2^{1/\delta}} \frac{e_i + 1}{p_i^{e_i \delta}} \leq \prod_{p \leq 2^{1/\delta}} e^{1/(\delta \log 2)} \leq e^{2^{1/\delta}/(\delta \log 2)}.$$

Equivalentemente,

$$\log d(n) - \delta \log n \leq \frac{2^{1/\delta}}{\delta \log 2}.$$

Ahora tomamos, concretamente,

$$\delta = \frac{(1 + \epsilon/2) \log 2}{\log \log n},$$

con lo que

$$2^{1/\delta} = 2^{\frac{\log \log n}{(1 + \epsilon/2) \log 2}} = e^{\frac{\log \log n}{1 + \epsilon/2}} = (\log n)^{1/(1 + \epsilon/2)}.$$

Por lo tanto,

$$\log d(n) \leq \frac{(1 + \epsilon/2) \log 2 \log n}{\log \log n} + \frac{(\log n)^{1/(1 + \epsilon/2)} \log \log n}{(1 + \epsilon/2) \log^2 2}.$$

Observemos ahora que, para todo n suficientemente grande,

$$\frac{(\log n)^{1/(1 + \epsilon/2)} \log \log n}{(1 + \epsilon/2) \log^2 2} \leq \frac{\epsilon \log 2 \log n}{2 \log \log n},$$

pues esto equivale a

$$\frac{\log^2 \log n}{(\log n)^{\frac{\epsilon/2}{1 + \epsilon/2}}} \leq \frac{\epsilon}{2} \left(1 + \frac{\epsilon}{2}\right) \log^3 2 \quad (1.7)$$

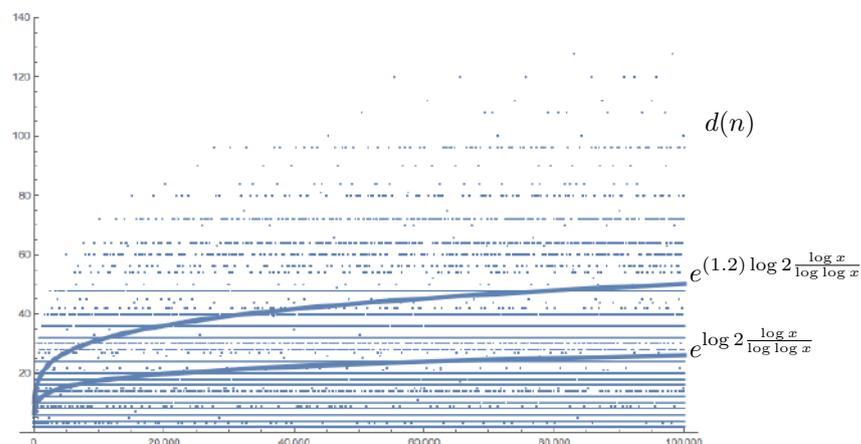
y el miembro izquierdo tiende a 0 con n (pues, si llamamos $x = \log n$, es de la forma $\frac{\log^2 x}{x^\alpha}$). Así pues,

$$\log d(n) \leq \frac{(1 + \epsilon/2) \log 2 \log n}{\log \log n} + \frac{(\epsilon/2) \log 2 \log n}{\log \log n} = \frac{(1 + \epsilon) \log 2 \log n}{\log \log n}.$$

■

Ahora bien, a la vista de la figura siguiente, lo primero que cabe pensar es que el teorema es falso, pues deberíamos ver que, a partir de cierto valor, la función $d(n)$ queda por debajo de la mayor de las dos curvas. Si extendemos la gráfica hasta, digamos, $n = 10^6$, la imagen es similar.

La razón es el fenómeno que hemos discutido antes del teorema. Hemos probado que la función $d(n)$ debe quedar por debajo de la curva mayor para todo n suficientemente grande, pero ¿cómo de grande?



El único punto de la prueba que requiere que n sea grande es (1.7) que, para $\epsilon = 0.2$, es, concretamente

$$\frac{\log^2 \log n}{(\log n)^{0.09}} \leq 0.037.$$

Ahora bien, llamando $x = \log n$, hemos visto que el máximo de la función

$$\frac{\log^2 x}{x^{0.09}}$$

se encuentra en $x^* = e^{2/0.09} \approx 3.58 \cdot 10^9$, luego para que el miembro izquierdo de la desigualdad anterior quede por debajo de 0.037, como mínimo $\log n$ debe superar dicho máximo, lo que requiere que $n \geq e^{3.58 \cdot 10^9} = 1.63 \cdot 10^{1554774245}$. De hecho, la función no desciende hasta 0.037 a partir de $x = 2.6 \cdot 10^{63}$, lo que exige que $n \geq e^{2.6 \cdot 10^{63}}$, y esto excede la capacidad de cómputo de un ordenador típico. En realidad nada impide que el teorema se cumpla para valores menores de n , pero sólo hemos probado que se cumple a partir de este valor.

En resumen: hemos razonado que los números naturales tienen una propiedad que nunca podremos ver que se cumple.

1.6 Continuidad de la derivada

Hemos visto que toda función derivable es continua, pero la derivada a su vez no tiene por qué ser continua.

Ejemplo La función

$$f(x) = \begin{cases} x^2 \operatorname{sen}(1/x) & \text{si } x \neq 0, \\ 0 & \text{si } x = 0, \end{cases}$$

es derivable en todo \mathbb{R} . En efecto, es derivable en $x = 0$ porque la derivada es

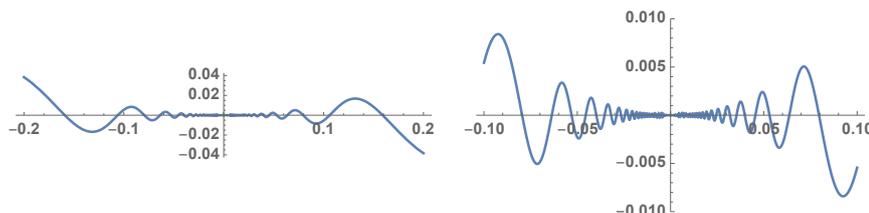
$$f'(0) = \lim_{h \rightarrow 0} \frac{h^2 \operatorname{sen}(1/h)}{h} = \lim_{h \rightarrow 0} h \operatorname{sen} \frac{1}{h} = 0,$$

y en los puntos restantes podemos aplicar las reglas de derivación para calcular

$$f'(x) = 2x \operatorname{sen}(1/x) + x^2 \cos(1/x)(-1/x^2) = 2x \operatorname{sen}(1/x) - \cos(1/x).$$

Puesto que $\lim_{x \rightarrow 0} 2x \operatorname{sen}(1/x) = 0$, es claro que no existe $\lim_{x \rightarrow 0} f'(x)$, luego $f'(x)$ es discontinua en $x = 0$.

He aquí dos imágenes de la gráfica de $f(x)$:



En la de la izquierda los dos ejes tienen la misma escala, mientras que en la de la derecha hemos hecho “trampa” y hemos graduado el eje vertical con una escala 50 veces mayor. Si ampliamos suficientemente usando la misma escala, la gráfica acaba siendo indistinguible del eje x . La discontinuidad de la derivada se debe a que la gráfica se confunde con una recta porque las oscilaciones se ven cada vez menores a medida que ampliamos la gráfica, pero no dejan de estar ahí, por lo que cerca de 0 hay puntos con cualquier pendiente entre -1 y 1 . ■

Sin embargo, las funciones discontinuas que son la derivada de otra función son relativamente escasas. Una muestra de ello nos la da el teorema siguiente:

Teorema 1.29 *Sea $f :]a, b[\rightarrow \mathbb{R}$ una función continua en $]a, b[$ y derivable en $]a, b[$ salvo a lo sumo en $x = c$. Si existe $\lim_{x \rightarrow c} f'(x)$, entonces f es derivable en c y $f'(c) = \lim_{x \rightarrow c} f'(x)$. En particular, f' es $x \rightarrow c$ continua en c .*

DEMOSTRACIÓN: Por la regla de L'Hôpital:⁷

$$f'(c) = \lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \rightarrow 0} \frac{f'(c+h)}{1} = \lim_{x \rightarrow c} f'(x).$$

■

Así pues, la derivada de una función no tiene por qué ser continua, pero si no es continua en un punto, podemos asegurar que el problema no será meramente que el límite en el punto no coincide con el valor de la derivada, sino que dicho límite no puede existir, tal y como hemos visto que sucede en el ejemplo de derivada discontinua que hemos dado. Más aún, las derivadas comparten una propiedad fundamental con las funciones continuas:

Teorema 1.30 (Darboux) *Si $f :]a, b[\rightarrow \mathbb{R}$ es una función derivable y tomamos números $a < u < v < b$, para todo número real y comprendido entre $f'(u)$ y $f'(v)$ existe $u \leq x \leq v$ tal que $f'(x) = y$.*

⁷La continuidad de f es necesaria para asegurar que el numerador tiende a 0, pues de lo contrario no se puede aplicar la regla.

DEMOSTRACIÓN: Supongamos que $f'(u) \leq y \leq f'(v)$. Si se dan las desigualdades opuestas se razona análogamente. Si se da una igualdad, basta tomar $x = u$ o $x = v$, según el caso, así que podemos suponer que $f'(u) < y < f'(v)$. Sea $g(t) = f(t) - yt$. Entonces g es continua en el intervalo $[u, v]$, luego alcanza su valor mínimo en un punto $u \leq x \leq v$. Ahora bien, $g'(u) = f'(u) - y < 0$, luego g no puede tener su valor mínimo en u (pues podemos tomar $h > 0$ suficientemente pequeño como para que $u < u + h < v$ y $f(u + h) < f(u)$). Igualmente $g'(v) = f'(v) - y > 0$, luego tampoco puede tener su valor mínimo en v . Así pues, $u < x < v$ y por el teorema de Fermat $g'(x) = f'(x) - y = 0$. ■

A su vez esto impone una limitación mayor a las discontinuidades posibles de una derivada:

Teorema 1.31 Sea $f :]a, b[\rightarrow \mathbb{R}$ una función derivable en $]a, b[$, sea $a < c < b$ y supongamos que existen

$$L_- = \lim_{x \rightarrow c^-} f'(x), \quad L_+ = \lim_{x \rightarrow c^+} f'(x),$$

entonces $L_- = L_+$, luego $f'(x)$ es continua en c .

DEMOSTRACIÓN: Supongamos, por ejemplo, que $L_- < L_+$. Si se da la desigualdad opuesta se razona análogamente. Tomamos

$$\epsilon = \frac{L_+ - L_-}{3}$$

y sea $\delta > 0$ tal que si $c - \delta < x < c$ se cumpla $|f'(x) - L_-| < \epsilon$, y si $c < x < c + \delta$ se cumpla $|f'(x) - L_+| < \epsilon$. Tomemos $c - \delta < u < c < v < c + \delta$ y sea $y \neq f'(c)$ tal que

$$f'(u) < L_- + \epsilon < y < L_+ - \epsilon < f'(v).$$

Por el teorema anterior existe $u < x < v$ tal que $f'(x) = y$. Como $y \neq f'(c)$, tiene que ser $x \neq c$, luego $c - \delta < x < c$ o bien $c < x < c + \delta$. En el primer caso $|y - L_-| < \epsilon$, y en el segundo $|y - L_+| < \epsilon$, en contra de la elección de y . Por lo tanto, $f'(x)$ tiene límite en c y el teorema 1.29 implica que es continua. ■

Así pues, una derivada no puede tener los dos límites laterales finitos en un punto de discontinuidad.

1.7 Ecuaciones diferenciales I

Una *ecuación diferencial* es una ecuación que tiene como incógnita una función $y : I \rightarrow \mathbb{R}$, donde I es un intervalo abierto en \mathbb{R} en el cual es derivable, y en la cual aparecen la variable x , la incógnita $y(x)$ y sus derivadas $y'(x)$, $y''(x)$, etc. Si la mayor derivada que aparece es la derivada n -sima, se dice que la ecuación diferencial tiene *orden n* .

Por ejemplo, una ecuación diferencial sencilla de primer orden es

$$y' = ky.$$

En general, las ecuaciones diferenciales tienen infinitas soluciones. Por ejemplo, supongamos que una solución $y : I \rightarrow \mathbb{R}$ de esta ecuación cumple $y(x) > 0$ para todo x en el intervalo I . Entonces podemos considerar la función $\log y(x)$, que cumple

$$(\log y)' = \frac{y'}{y} = k.$$

Así pues, $\log y$ tiene la misma derivada que la función kx , luego el teorema 1.16 nos da que existe una constante c_0 tal que $\log y = kx + c_0$, luego $y = e^{c_0} e^{kx}$. Llamando $c = e^{c_0}$ concluimos que $y = ce^{kx}$.

Es claro que toda función de la forma $y = ce^{kx}$ es solución de la ecuación, y hemos probado que las únicas soluciones positivas de la ecuación diferencial dada son las de esta forma, para cada constante $c > 0$. Pero si y es solución con $y < 0$, entonces $-y$ es solución con $-y > 0$, de donde se sigue que $y = ce^{kx}$, para cierta constante $c < 0$. Con esto hemos probado parte del teorema siguiente:

Teorema 1.32 *Las únicas soluciones de la ecuación diferencial $y' = ky$, con $k \neq 0$ son las de la forma $y = ce^{kx}$, para cada constante $c \in \mathbb{R}$.*

Faltaría probar que si una solución se anula en un punto, entonces es idénticamente nula, pues eso implica a su vez que una solución no puede tomar a la vez valores positivos y negativos, ya que entonces, por ser continua, se anularía en algún punto. No obstante, no merece la pena entrar en ello, porque el teorema anterior será inmediato en cuanto hayamos demostrado el teorema siguiente, que no trata sobre ecuaciones diferenciales, sino sobre sistemas (lineales) de ecuaciones diferenciales:

Teorema 1.33 *Supongamos que $y_1, \dots, y_n, \bar{y}_1, \dots, \bar{y}_n : I \rightarrow \mathbb{R}$ son funciones derivables en el intervalo abierto I que satisfacen un sistema de ecuaciones diferenciales de la forma*

$$y_1' + a_{11}y_1 + \dots + a_{1n}y_n = b_1$$

.....

$$y_n' + a_{n1}y_1 + \dots + a_{nn}y_n = b_n$$

para ciertas funciones continuas $a_{ij}, b_i : I \rightarrow \mathbb{R}$, así como que existe un $x_0 \in I$ tal que $y_i(x_0) = \bar{y}_i(x_0)$ para $i = 1, \dots, n$. Entonces $y_i = \bar{y}_i$ para $i = 1, \dots, n$.

En otras palabras, un sistema de ecuaciones diferenciales de este tipo tiene a lo sumo una solución (y_1, \dots, y_n) que cumpla unas condiciones iniciales dadas de la forma

$$y_1(x_0) = u_0, \quad \dots, \quad y_n(x_0) = u_n.$$

DEMOSTRACIÓN: Tenemos que

$$\begin{array}{ll} y'_1 + a_{11}y_1 + \cdots + a_{1n}y_n = b_1 & \bar{y}'_1 + a_{11}\bar{y}_1 + \cdots + a_{1n}\bar{y}_n = b_1 \\ \dots\dots\dots & \dots\dots\dots \\ y'_n + a_{n1}y_1 + \cdots + a_{nn}y_n = b_n & \bar{y}'_n + a_{n1}\bar{y}_1 + \cdots + a_{nn}\bar{y}_n = b_n \end{array}$$

luego, si llamamos $z_i = y_i - \bar{y}_i$, restando las ecuaciones correspondientes de cada sistema resulta que las funciones z_i cumplen:

$$\begin{array}{l} z'_1 + a_{11}z_1 + \cdots + a_{1n}z_n = 0 \\ \dots\dots\dots \\ z'_n + a_{n1}z_1 + \cdots + a_{nn}z_n = 0 \end{array}$$

Y además $z_i(x_0) = 0$ para $i = 1, \dots, n$. En estas condiciones, tenemos que demostrar que $z_i = 0$ para $i = 1, \dots, n$, pues esto implica que $y_i = \bar{y}_i$. Para ello consideramos la función $E(x) = z_1^2(x) + \cdots + z_n^2(x) \geq 0$.

Sabemos que $E(x_0) = 0$ y basta probar que $E = 0$. Supongamos, por el contrario, que existe un punto $x_1 \in I$ tal que $E(x_1) > 0$. Sean $p, q \in I$ tales que $p < x_0, x_1 < q$. Así $J = [p, q] \subset I$ y por el teorema de Weierstrass [ITAN 3.31], las funciones continuas a_{ij} están acotadas en J , es decir, existe una constante $K > 0$ tal que $|a_{ij}(x)| \leq K$, para todo $x \in J$. Seguidamente observamos que

$$E' = \sum_{i=1}^n 2z_i z'_i = - \sum_{i=1}^n \sum_{j=1}^n a_{ij} 2z_i z_j.$$

Ahora usamos que $|2ab| \leq a^2 + b^2$ (basta desarrollar $(a + b)^2 \geq 0$ y $(a - b)^2 \geq 0$) para concluir que $-(a^2 + b^2) \leq 2ab \leq a^2 + b^2$.

Así, para $x \in]p, q[$,

$$|E'| \leq \sum_{i,j=1}^n |a_{ij}| |2z_i z_j| \leq \sum_{i,j=1}^n K(z_i^2 + z_j^2) \leq \sum_{i,j=1}^n 2KE = 2n^2 KE = ME,$$

luego $-ME \leq E' \leq ME$, para todo $x \in]p, q[$.

Si $x_0 < x_1$, observamos que

$$\frac{d}{dx}(e^{-Mx} E) = e^{-Mx}(E' - ME) \leq 0,$$

luego la función $e^{-Mx} E$ es decreciente en $]p, q[$, lo cual nos da una contradicción, pues entonces $e^{-Mx_1} E(x_1) \leq e^{-Mx_0} E(x_0) = 0$, luego $E(x_1) \leq 0$.

Similarmente, si $x_1 < x_0$ razonamos que la función $e^{Mx} E$ es creciente y llegamos a que $e^{Mx_1} E(x_1) \leq e^{Mx_0} E(x_0) = 0$ y nuevamente $E(x_1) \leq 0$. ■

Ahora el teorema 1.32 es inmediato (incluso sin la discusión previa), pues es evidente que las funciones $y = ce^x$ son soluciones de la ecuación diferencial (que es de la forma considerada en el teorema anterior, es decir, $y' - ky = 0$) y cumplen $y(0) = c$, luego si y es una solución arbitraria y $c = y(0)$, entonces $\bar{y} = ce^{kx}$ es otra solución con $y(0) = \bar{y}(0)$, luego el teorema nos da que $y = \bar{y}$.

Ejemplo: Desintegración radiactiva Los elementos radiactivos son aquellos cuyos átomos, al cabo de un tiempo, se desintegran emitiendo radiación y transformándose, bien en otros elementos, o bien en isótopos del mismo elemento. Por ejemplo, el uranio 288 se desintegra en torio 234, el cual a su vez se desintegra en protactinio 234, que a su vez se desintegra en uranio 234, y esto sólo es el principio de una larga cadena que termina en el plomo, que es estable.

Lo más curioso de la desintegración radiactiva es que es imposible predecir cuándo se desintegrará un átomo dado, pero, al mismo tiempo, el proceso sigue una ley estadística muy simple, pues la proporción de átomos de una muestra de un elemento que se desintegra por unidad de tiempo es una constante que depende únicamente del elemento en cuestión. Matemáticamente, si llamamos $N(t)$ a la cantidad de átomos radiactivos de una muestra (ya sea el número de átomos propiamente dicho, o su masa, o cualquier otra magnitud proporcional al número de átomos), se cumple que

$$N' = -\lambda N,$$

donde la constante $\lambda > 0$ recibe el nombre de *constante de desintegración radiactiva*. En virtud del teorema 1.32 sabemos que

$$N = N_0 e^{-\lambda t},$$

que es ciertamente la solución particular que cumple $N(0) = N_0$. A menudo, en lugar de λ se usa el *periodo de semidesintegración*, que es el tiempo que tiene que transcurrir para que la muestra se reduzca a la mitad. Es fácil encontrar su relación con λ : se trata del tiempo t_0 que cumple

$$N_0 e^{-\lambda(t+t_0)} = \frac{N_0}{2} e^{-\lambda t},$$

que equivale a $e^{-\lambda t_0} = 1/2$, luego $-\lambda t_0 = -\log 2$, luego $t_0 = \log 2/\lambda$. En términos de t_0 la ecuación de decaimiento es

$$N = N_0 e^{-\log 2t/t_0} = N_0 2^{-t/t_0}.$$

Por ejemplo, el carbono tiene un isótopo radiactivo, el carbono-14, que se desintegra en el carbono ordinario (carbono-12) con un periodo de semidesintegración de 5730 años. El carbono-14 se encuentra en la atmósfera con relativa abundancia, y las plantas lo incorporan constantemente a su materia orgánica a través de la fotosíntesis, y a su vez los animales lo incorporan a través de su alimentación, de modo que la cantidad de carbono-14 en un ser vivo permanece más o menos constante a lo largo de su vida, pero tras su muerte ya no incorporan más carbono-14, por lo que la cantidad residual puede usarse para determinar aproximadamente la antigüedad de unos restos orgánicos.

En la práctica no es fácil contar cuántos átomos de carbono-14 contiene una muestra, pero sí es fácil contar los átomos que se desintegran, porque cada desintegración emite radiación fácilmente detectable. Consideremos un caso concreto:

En un trozo de madera “fresca”, la tasa de decaimiento del carbono-14 es de 13.6 átomos por minuto y por gramo, mientras que en un resto antiguo se cuentan 6 átomos por minuto y por gramo. Determinar la antigüedad de dicho trozo.

Tenemos que, en la actualidad, el número N^* de átomos de carbono-14 de un gramo de madera fresca cumple

$$-\lambda N^* = \frac{dN^*}{dt} = 13.6,$$

mientras que el número N de átomos en un gramo de la muestra cumple

$$-\lambda N = \frac{dN}{dt} = 6,$$

de donde deducimos que $N/N^* = 6/13.6 = 0.441176$. Si suponemos que el número N^* de átomos por gramo en la muestra fresca es aproximadamente igual que el número N_0 de átomos por gramo que tenía la muestra antigua en el momento en el que fue talada de su árbol (esta hipótesis no siempre es aceptable, pero se sabe cómo corregirla) concluimos que

$$2^{-t/5730} = N/N_0 = 0.441176,$$

luego $-(t/5730) \log 2 = \log 0.441176$ y llegamos a que la edad de la muestra es de aproximadamente

$$t = -5730 \frac{\log 0.441176}{\log 2} = 6765 \text{ años.} \quad \blacksquare$$

El teorema 1.33 se generaliza trivialmente a sistemas de ecuaciones de cualquier orden. Vamos a formular el caso de orden 2:

Teorema 1.34 *Supongamos que $y_1, \dots, y_n, \bar{y}_1, \dots, \bar{y}_n : I \rightarrow \mathbb{R}$ son funciones derivables en el intervalo abierto I que satisfacen un sistema de ecuaciones diferenciales de la forma*

$$y_1'' + a_{11}y_1' + \dots + a_{1n}y_n' + b_{11}y_1 + \dots + b_{1n}y_n = c_1$$

.....

$$y_n'' + a_{n1}y_1' + \dots + a_{nn}y_n' + b_{n1}y_1 + \dots + b_{nn}y_n = c_n$$

para ciertas funciones continuas $a_{ij}, b_{ij}, c_i : I \rightarrow \mathbb{R}$, así como que existe un $x_0 \in I$ tal que $y_i(x_0) = \bar{y}_i(x_0)$, $y_i'(x_0) = \bar{y}_i'(x_0)$ para $i = 1, \dots, n$. Entonces $y_i = \bar{y}_i$ para $i = 1, \dots, n$.

DEMOSTRACIÓN: Basta definir $z_i = y_i'$, $\bar{z}_i = \bar{y}_i'$, y entonces tanto las funciones $y_1, \dots, y_n, z_1, \dots, z_n$ como $\bar{y}_1, \dots, \bar{y}_n, \bar{z}_1, \dots, \bar{z}_n$ satisfacen el sistema de

todas las soluciones están definidas en \mathbb{R} , luego están determinadas también por los valores de $y(0)$ e $y'(0)$, luego cualquier solución puede expresarse en la forma

$$y = A \cos kx + B \operatorname{sen} kx.$$

Sólo falta probar que la segunda expresión recorre también todas las soluciones, pero ello se debe a que $y = C \cos(kx + \phi)$ cumple ciertamente la ecuación diferencial y además $y(0) = C \cos \phi$, $y'(0) = -Ck \operatorname{sen} \phi$, luego basta elegir C y ϕ para que se cumpla

$$(u, -v/k) = C(\cos \phi, \operatorname{sen} \phi)$$

y entonces la función $y(x)$ correspondiente cumple $y(0) = u$, $y'(0) = v$, por lo que las soluciones de esta forma recorren todas las soluciones posibles. ■

Si una magnitud $y(t)$ evoluciona en el tiempo satisfaciendo la ecuación diferencial $y'' = -ky$, se dice que es un *oscilador armónico*. Según el teorema anterior, $y = C \cos(kt + \phi)$ oscila periódicamente entre las *amplitudes* $\pm C$ con un *periodo* $T = 2\pi/k$, es decir, que vuelve al mismo estado al cabo de T unidades de tiempo, y con una *frecuencia* $\nu = 1/T = k/2\pi$ (que es el número de oscilaciones por unidad de tiempo).

Ecuaciones lineales homogéneas de segundo orden con coeficientes constantes Vamos a ver cómo resolver cualquier ecuación de este tipo, donde “homogénea” significa que el término independiente es nulo. Consideramos, pues, una ecuación de la forma

$$z''(t) + 2a_1 z'(t) + a_0 z(t) = 0,$$

donde los coeficientes a_1 y a_0 son constantes. (Obviamente el 2 en el coeficiente de z' es irrelevante, y lo hemos incluido porque así la solución se expresa de forma ligeramente más simple.)

Para resolverla vamos a usar un “truco”, y es que vamos a suponer que los coeficientes $2a_1 = u + vi$, $a_0 = r + si$ son números complejos, al igual que la función $z(t) = x(t) + y(t)i$. Esto tiene sentido y no se sale de la teoría que hemos desarrollado, pues si sustituimos en la ecuación y separamos la parte real de la imaginaria lo que tenemos es el sistema

$$x''(t) = -ux'(t) + vy'(t) - rx(t) + sy(t),$$

$$y''(t) = -vx'(t) - uy'(t) - sx(t) - ry(t),$$

y el teorema 1.34 nos garantiza que este sistema de ecuaciones tiene a lo sumo una solución para cada conjunto de condiciones iniciales

$$z(t_0) = x_0 + y_0 i, \quad z'(t_0) = x'_0 + y'_0 i$$

(es decir, para las condiciones $z(t_0) = z_0$, $z'(t_0) = z'_0$).

Sólo necesitamos observar que cuando consideramos a una función vectorial $z(t) = (x(t), y(t))$ como función compleja $z(t) = x(t) + y(t)i$ se siguen cumpliendo las propiedades siguientes:

- $(z_1 + z_2)'(t) = z_1'(t) + z_2'(t)$.
- $(az)'(t) = az'(t)$ (para $a \in \mathbb{C}$).
- $(e^{at})' = ae^{at}$ (para $a \in \mathbb{C}$).

Las dos primeras propiedades son inmediatas y la segunda se cumple porque, de hecho, la función e^{at} está definida para todo $t \in \mathbb{C}$ y es derivable en todo el plano complejo, y hemos visto que las funciones complejas cumplen las mismas reglas de derivación que las reales.

Consideremos una posible solución de la forma $z(t) = e^{kt}$, con $k \in \mathbb{C}$. Se comprueba inmediatamente que satisface la ecuación diferencial si y sólo si k cumple la ecuación

$$k^2 + 2a_1k + a_0 = 0,$$

es decir, si $k = -a_1 \pm d$, donde $d = \sqrt{a_1^2 - a_0}$. (Observemos que d puede ser un número complejo aunque partamos de coeficientes reales, y ésta es la razón por la que hemos pasado al caso complejo.) Supongamos en primer lugar que $d \neq 0$, con lo cual hemos encontrado dos soluciones: $z = e^{-a_1 t} e^{\pm dt}$. Ahora bien, es fácil ver que cualquier función de la forma

$$z = e^{-a_1 t} (Ae^{dt} + Be^{-dt}), \quad A, B \in \mathbb{C}$$

es también solución. Vamos a probar que éstas son todas las soluciones de la ecuación, lo cual equivale a probar que, para cada par de condiciones iniciales $z_0, z_0' \in \mathbb{C}$ es posible encontrar A y B tales que la solución correspondiente las satisfaga. No perdemos generalidad⁸ si suponemos que $t_0 = 0$, en cuyo caso las condiciones iniciales son

$$A + B = z_0, \quad -a_1(A + B) + d(A - B) = z_0'.$$

Es claro que este sistema de ecuaciones tiene solución para A y B , pero en lugar de calcularla es más práctico expresar la solución en términos de $A + B$ y $A - B$, así:

$$z(t) = e^{-a_1 t} \left((A + B) \frac{e^{ct} + e^{-ct}}{2} + (A - B) \frac{e^{ct} - e^{-ct}}{2} \right),$$

con lo que

$$z(t) = e^{-a_1 t} \left(z_0 \cosh dt + \left(\frac{a_1 z_0}{d} + \frac{z_0'}{d} \right) \sinh dt \right) \quad (1.8)$$

es una solución que cumple $z(0) = z_0$, $z'(0) = z_0'$. Una expresión alternativa para la solución se obtiene llamando $q = \sqrt{a_0 - a_1^2}$, con lo que $d = iq$, y entonces las relaciones

$$\operatorname{sen} z = \frac{1}{i} \operatorname{senh} iz, \quad \operatorname{cos} z = \operatorname{cosh} iz,$$

⁸Si z es una solución definida en un intervalo I que contenga a t_0 , entonces $z(t - t_0)$ es también solución de la ecuación diferencial y está definida en 0, luego, según vamos a probar, será de la forma indicada para ciertos A y B , pero es fácil ver entonces que z también lo es, para otros valores de A y B .

nos dan que

$$z(t) = e^{-a_1 t} \left(z_0 \cos qt + \left(\frac{a_1 z_0}{q} + \frac{z'_0}{q} \right) \operatorname{sen} qt \right). \quad (1.9)$$

De este modo, si los coeficientes de la ecuación de partida son reales, al igual que las condiciones iniciales, tenemos que o bien (1.8) o bien (1.9) muestra que la solución que hemos obtenido es real, según que sea $a_1^2 > a_0$ (en cuyo caso $d \in \mathbb{R}$) o bien $a_1^2 < a_0$ (en cuyo caso $q \in \mathbb{R}$).

Nos falta considerar el caso en que $d = 0$, en el que sólo tenemos una solución $z = e^{-a_1 t}$, pero en este caso la ecuación se reduce a

$$z''(t) + 2a_1 z'(t) + a_1^2 z(t) = 0,$$

y una comprobación rutinaria muestra que otra solución es $z(t) = te^{-a_1 t}$. A partir de ahí, el razonamiento precedente se adapta para justificar que la solución general es

$$z(t) = e^{-a_1 t} (A + Bt)$$

o, en términos de las condiciones iniciales:

$$z(t) = e^{-a_1 t} (z_0 + (a_1 z_0 + z'_0)t).$$

Así hemos probado el teorema siguiente:

Teorema 1.36 *La solución $z : \mathbb{C} \rightarrow \mathbb{C}$ de una ecuación diferencial de la forma*

$$z''(t) + 2a_1 z'(t) + a_0 z(t) = 0, \quad z(0) = z_0, \quad z'(0) = z'_0,$$

donde a_0, a_1 son números complejos con $d = \sqrt{a_1^2 - a_0} \neq 0$, es la función

$$z(t) = e^{-a_1 t} \left(z_0 \cosh dt + \left(\frac{a_1 z_0}{d} + \frac{z'_0}{d} \right) \operatorname{senh} dt \right)$$

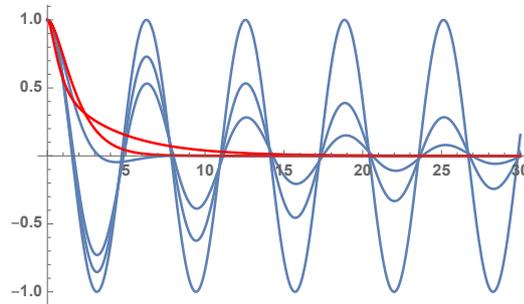
o, equivalentemente, haciendo $d = iq$,

$$z(t) = e^{-a_1 t} \left(z_0 \cos qt + \left(\frac{a_1 z_0}{q} + \frac{z'_0}{q} \right) \operatorname{sen} qt \right).$$

Si $d = 0$ la solución es

$$z(t) = e^{-a_1 t} (z_0 + (a_1 z_0 + z'_0)t).$$

La figura siguiente muestra varios ejemplos de soluciones con $a_0 = 1$, $z_0 = 1$, $z'_0 = 0$ para varios valores de a_1 :



Cuando $a_1 = 0$ tenemos una oscilación armónica, pues la ecuación es del tipo considerado en el teorema 1.35. A medida que a_1 va aumentando, su efecto es el de atenuar la amplitud de oscilaciones (pero no su frecuencia), que se reduce más rápidamente cuanto mayor es $a_1 < a_0^2$. Cuando $a_1 = a_0^2$ desaparecen las oscilaciones y la función $y(t)$ tiende a 0 sin hacerse nunca negativa. Para valores mayores de a_1 el comportamiento es el mismo, sólo que el decrecimiento es cada vez más lento.

Una caracterización de las cónicas Recordemos de [IGE 7.1] que si d es una recta y F un punto exterior, la cónica de directriz d y foco F está formada por los puntos tales que la razón entre las distancias a F y a d es constante (y recibe el nombre de *excentricidad* de la cónica). Toda cónica que no sea una circunferencia es de esta forma.

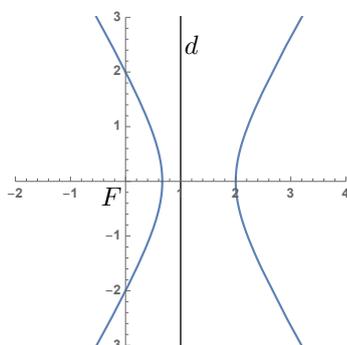
Si tomamos un sistema de referencia con origen en el foco de modo que la directriz sea la recta vertical $x = p > 0$, entonces la distancia a la directriz de un punto de coordenadas polares (ρ, θ) es $|p - \rho \cos \theta|$, luego la ecuación de la cónica de excentricidad ϵ es

$$\frac{\rho}{|p - \rho \cos \theta|} = \epsilon.$$

Si $\epsilon \leq 1$ se cumple que $p - \rho \cos \theta \geq 0$, pues para que fuera negativo tendría que ser $\rho > p$ y entonces $0 < \rho \cos \theta - p < \rho$, con lo que el cociente con el valor absoluto sería > 1 . Por lo tanto, si quitamos el valor absoluto y pasamos a

$$\frac{\rho}{p - \rho \cos \theta} = \epsilon, \quad (1.10)$$

sólo perdemos puntos cuando la excentricidad $\epsilon > 1$, es decir, cuando la cónica es una hipérbola. En tal caso, los puntos que cumplían la ecuación original con denominador negativo ya no cumplen la nueva ecuación, pues $\rho/\epsilon > 0$. Concretamente, hemos quitado los puntos con $x = \rho \cos \theta > p$, es decir, los puntos que están a la izquierda de la directriz $x = p$, que constituyen una de las ramas de la hipérbola, la que rodea al segundo foco.



Así pues, la ecuación (1.10) la satisfacen los puntos de la cónica de excentricidad $\epsilon > 0$, foco $(0, 0)$ y directriz $x = p$, salvo por el hecho de que si se trata de una parábola estamos excluyendo la rama que no rodea al origen.

Despejando ρ y llamando $l = p\epsilon$ queda

$$\rho = \frac{l}{1 + \epsilon \cos \theta},$$

Cualquier curva que satisfaga una ecuación de esta forma con $l, \epsilon > 0$ (entendiendo que a los ángulos θ que hagan el denominador ≤ 0 no les corresponde ningún punto de la curva) es una cónica de excentricidad $\epsilon > 0$, pero si admitimos $\epsilon = 0$ queda $\rho = l$, con lo que hemos incorporado a las circunferencias como cónicas de excentricidad 0. Un poco más en general, si incorporamos una nueva constante:

$$\rho = \frac{l}{1 + \epsilon \cos(\theta + \phi)}, \quad (1.11)$$

su efecto es simplemente girar la cónica un ángulo ϕ respecto al origen, luego las curvas con una ecuación de este tipo son todas las cónicas de excentricidad $\epsilon \geq 0$ con un foco en el origen de coordenadas, sin la restricción de que su directriz (en el caso $\epsilon > 0$) tenga que ser una recta vertical $x = p$.

Vamos a caracterizar estas ecuaciones en términos de una ecuación diferencial. Para ello llamamos $u = 1/\rho$, $u_0 = 1/k > 0$ y $C = \epsilon/l \geq 0$, de modo que

$$u(\theta) = \frac{1}{l}(1 + \epsilon \cos(\theta + \phi)) = u_0 + C \cos(\theta + \phi).$$

Ahora observamos que $u'' = -C \cos(\theta + \phi) = u_0 - u$, luego tenemos que las funciones $u(\theta)$ son soluciones de la ecuación diferencial

$$u'' + u = u_0, \quad u_0 > 0. \quad (1.12)$$

Vamos a comprobar que no hay más. Para ello observamos que si $u(\theta)$ es cualquier solución de esta ecuación diferencial, la función $v = u - u_0$ cumple la ecuación $v'' = -v$, que es de la forma considerada en el teorema 1.35 con $k = 1$, luego sus soluciones son, ni más ni menos, las funciones de la forma

$$v = C \cos(\theta + \phi),$$

para cualquier par de constantes C y ϕ . Ahora bien, si $C < 0$, la misma función puede expresarse como

$$v = -C \cos(\theta + \phi + \pi),$$

luego cambiando si es preciso C por $-C$ y ϕ por $\phi + \pi$ resulta que no perdemos generalidad si suponemos que $C \geq 0$. Así pues, las soluciones de (1.12) son las de la forma

$$u = u_0 + C \cos(\theta + \phi),$$

para cualquier $C \geq 0$ y cualquier ϕ , y sus inversas, $\rho = 1/u$ son precisamente las funciones (1.11), con $l = 1/u_0$ y $\epsilon = lC \geq 0$.

En otros términos, hemos demostrado que las soluciones de la ecuación diferencial (1.12) son las funciones $u(\theta)$ cuyas inversas $\rho = 1/u$ se corresponden con las ecuaciones en coordenadas polares (1.11) de las cónicas que tienen un foco en el origen de coordenadas, con $l = 1/u_0$.

Más aún, si ρ es cualquier función no negativa y $u = \rho^{-1}$, derivando obtenemos

$$u' = -\rho^{-2}\rho', \quad u'' = -\rho^{-3}\rho'^2 - \rho^{-2}\rho'',$$

luego

$$u'' + u = -\frac{\rho''}{\rho^2} + \frac{2\rho'^2}{\rho^3} + \frac{1}{\rho}, \quad (1.13)$$

luego u satisface (1.12) si y sólo si ρ satisface la ecuación diferencial

$$\frac{\rho''}{\rho^2} - \frac{2\rho'^2}{\rho^3} - \frac{1}{\rho} = u_0.$$

Con esto hemos probado el teorema siguiente:

Teorema 1.37 *Las soluciones de la ecuación diferencial*

$$\frac{\rho''}{\rho^2} - \frac{2\rho'^2}{\rho^3} - \frac{1}{\rho} = k,$$

con $k > 0$, son las funciones de la forma

$$\rho = \frac{l}{1 + \epsilon \cos(\theta + \phi)},$$

donde $l = 1/k$, para ciertas constantes $\epsilon \geq 0$ y ϕ , que son las ecuaciones en coordenadas polares de todas las cónicas con un foco en el origen y excentricidad ϵ .

1.8 La diferencial de una función

Según hemos visto, la función $y = x^3$ cumple

$$\frac{dy}{dx} = 3x^2.$$

Aunque hemos introducido esta notación, debida a Leibniz, en todo momento hemos usado la notación alternativa de Euler: $y' = 3x^2$. Las únicas excepciones las hemos hecho a la hora de enunciar la regla de la cadena y el teorema de la función inversa, donde la notación de Leibniz resulta especialmente sugerente. Eliminando precisiones que se pueden sobrentender por el contexto se reducen a:

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}, \quad \frac{dx}{dy} = \frac{1}{dy/dx}.$$

Estas fórmulas sugieren que una derivada dy/dx es una especie de cociente. El hecho es que no hemos definido las derivadas como cocientes ni hemos demostrado las fórmulas anteriores suponiendo que las derivadas sean ninguna clase de cociente, pero no es menos cierto que “se comportan como si fueran cocientes”. Más aún, Leibniz consideraba las derivadas “como si fueran cocientes”. Un razonamiento “típico” en los inicios del “cálculo diferencial” para calcular la derivada de $y = x^3$ era como sigue:

Fijemos un punto x y consideremos una cantidad no nula pero infinitamente pequeña dx (léase “diferencial de x ”). Al pasar de x a $x + dx$, la función y experimenta también un incremento infinitesimal dy , dado por

$$dy = (x + dx)^3 - x^3 = 3x^2 dx + 3x dx^2 + dx^3$$

de modo que el cociente del incremento infinitesimal que experimenta y sobre el incremento infinitesimal experimentado por x es

$$\frac{dy}{dx} = 3x^2 + 3x dx + dx^2.$$

Vemos que éste consta de una parte “apreciable” $3x^2$ y de una parte infinitesimal inapreciable $3x dx + dx^2$. En efecto, el hecho de que dx no sea meramente un incremento pequeño, sino “infinitamente pequeño” conlleva que, por grande que pueda ser x , el producto $x dx$ es infinitamente pequeño, y con mayor razón lo será el producto dx^2 de dos cantidades infinitamente pequeñas. Por consiguiente, podemos afirmar que “en la practica”:

$$\frac{dy}{dx} = 3x^2,$$

hecho que podía expresarse equivalentemente como

$$dy = 3x^2 dx.$$

Esto de considerar que dx es una cantidad no nula (para que tenga sentido el cociente dy/dx) y a la vez que al final podamos desembarazarnos de ella como si fuera nula fue objeto de muchas críticas que ponían en cuestión los fundamentos del “cálculo infinitesimal” a la que los matemáticos no sabían responder más que con la evidencia de que funcionaba y daba respuestas a numerosos problemas cuya corrección podía ser comprobada, aunque los medios que conducían a ellas pudieran ser cuestionable.

Estos problemas de fundamentación desaparecieron cuando los matemáticos pudieron justificar el cálculo infinitesimal definiendo las derivadas dy/dx , no como cocientes de cantidades infinitesimales, sino como límites de cocientes de incrementos finitos, que es lo que hemos hecho nosotros. Ahora bien, la definición de derivada que hemos dado trata a dy/dx como una mera notación equivalente a y' , de modo que no justifica ecuaciones como $dy = 3x^2 dx$, en los que dy y dx tienen “vida independiente”. Sin embargo, sucede que este tipo de ecuaciones resultan útiles en muchos contextos y pueden ser justificadas sin

necesidad de apelar a misteriosas cantidades no nulas infinitamente pequeñas, mediante las definiciones oportunas.⁹

Para ello, si $f :]a, b[\rightarrow \mathbb{R}$ es una función derivable en un punto x , llamaremos *diferencial* de f en el punto x a la aplicación $df|_x : \mathbb{R} \rightarrow \mathbb{R}$ dada por $df|_x(\Delta x) = f'(x) \Delta x$, donde, en principio, Δx (léase “incremento de x ”) no es más que una variable (la que hasta ahora hemos llamado usualmente h), a la que damos ese nombre porque pretende recorrer posibles incrementos de la variable x . En estos términos, sabemos que

$$f(x + \Delta x) - f(x) \approx df|_x(\Delta x),$$

donde la aproximación no es una aproximación cualquiera, sino que el miembro izquierdo es la variación la función f cuando pasamos de x a $x + \Delta x$ y el miembro derecho es la variación de la recta tangente a la gráfica de f en el punto x cuando pasamos de x a $x + \Delta x$, o también, $df|_x(\Delta x)$ es lo que habría variado f al pasar de x a $x + \Delta x$ si hubiera mantenido el ritmo de crecimiento que tiene en x .

Por ejemplo, si $y = x^3$, de acuerdo con las definiciones que hemos dado, podemos afirmar que

$$dy|_x(\Delta x) = 3x^2 \Delta x.$$

Ahora bien, si consideramos por otra parte la función x (es decir, la función $f(x) = x$), tenemos que su derivada vale 1 en cualquier punto, luego, por definición, $dx|_x(\Delta x) = \Delta x$, y así, podemos expresar la igualdad anterior en la forma

$$dy|_x(\Delta x) = 3x^2 dx|_x(\Delta x).$$

En otras palabras, para cada número real x tenemos definidas las funciones $dy|_x$ y $3x^2 dx|_x$, y acabamos de probar que son iguales, es decir, que

$$dy|_x = 3x^2 dx|_x$$

Finalmente, si consideramos las funciones que a cada x le hacen corresponder las funciones $dy|_x$ y $3x^2 dx|_x$, tenemos que ambas son iguales, y podemos expresar esta igualdad en la forma

$$dy = 3x^2 dx.$$

Definición 1.38 Si $f : I \rightarrow \mathbb{R}$ es una función derivable en un intervalo abierto $I \subset \mathbb{R}$, a cada $x \in I$ le podemos asociar las funciones $df|_x, dx|_x : \mathbb{R} \rightarrow \mathbb{R}$ dadas por

$$df|_x(\Delta x) = \left. \frac{df}{dx} \right|_x \Delta x, \quad dx|_x(\Delta x) = \Delta x,$$

⁹En 1961 Abraham Robinson descubrió una forma de definir con rigor matemático el concepto de “número real infinitamente pequeño”, de modo que los argumentos antiguos que involucraban cantidades infinitesimales podían formalizarse en muchos casos. Así surgió lo que se denomina *Análisis no estándar*. El enfoque de Robinson dependía enormemente de la teoría de modelos, y posteriormente surgieron otras variantes que requerían un menor conocimiento de la lógica matemática, pero no hay realmente ninguna presentación del análisis no estándar en la que no sea fácil perderse en contradicciones si no se está bien familiarizado con las sutilezas de la lógica formal que son irrelevantes en la práctica matemática habitual.

y la relación entre ambas es que

$$df|_x = \left. \frac{df}{dx} \right|_x dx|_x$$

para todo $x \in I$, luego podemos expresar esto en la forma

$$df = \frac{df}{dx} dx.$$

Así pues, con estas notaciones está justificado que las ecuaciones

$$\frac{dy}{dx} = 3x^2, \quad dy = 3x^2 dx$$

son equivalentes, y la segunda expresa que, para cualquier punto x , el incremento que experimenta la recta tangente a la gráfica de y en el punto $(x, y(x))$ cuando se aumenta x en una cantidad Δx es igual a $3x^2$ por el incremento que experimenta la recta tangente a la gráfica de x en el punto (x, x) , es decir, es igual a $3x^2 \Delta x$. En suma, la idea subyacente a esta notación es que dy no significa “incremento de la función y ” sino “incremento de la recta tangente a la gráfica de y ”.

No obstante, como el pensamiento es libre, dado que las definiciones precedentes nos legitiman a escribir con rigor igualdades como $dy = 3x^2 dx$, nada nos impide concebirlas como que $3x^2 dx$ es el incremento infinitesimal que experimenta la función $y = x^3$ cuando la variable x experimenta un incremento infinitesimal dx .

Capítulo II

El cálculo integral

Supongamos que un coche se mueve por una carretera en línea recta. En el instante $t_0 = 0$ se encuentra en la posición x_0 y en cada instante t se mueve a velocidad $v(t) = e^{t^2}$ m/s. ¿Dónde se encontrará en el instante $t = 2$?

Más precisamente, si llamamos $x(t)$ a la función que nos da la posición del coche en cada instante t , las hipótesis son que

$$\frac{dx}{dt} = e^{t^2}, \quad x(0) = x_0,$$

y con estos datos queremos calcular $x(2)$ o, más en general, la función $x(t)$.

La solución tiene que ser única por el teorema 1.16, que nos asegura que si dos funciones $x_1(t)$ y $x_2(t)$ cumplen la condición sobre la derivada, entonces tienen que cumplir $x_2(t) = x_1(t) + k$, para cierta constante k , la cual tiene que ser nula por la segunda condición, luego se trata de la misma función.

Si la velocidad hubiera sido $v(t) = 2te^{t^2}$, es fácil ver que la función requerida tiene que ser de la forma $x(t) = e^{t^2} + k$, donde a su vez podemos calcular k para que se cumpla $x(0) = x_0$, lo que obliga a que $x(t) = e^{t^2} - 1 + x_0$.

Cuando $v(t) = e^{t^2}$ no es fácil encontrar una función que cumpla lo requerido. No obstante, sí que es fácil calcularla para instantes concretos. Veamos, por ejemplo, cómo calcular $x(2)$.

Como primera aproximación, puesto que $v(0) = e^0 = 1$ m/s, podemos afirmar que, aproximadamente, en 2 segundos habrá recorrido unos 2 m, por lo que $x(2) \approx x_0 + 2$. Ahora bien, este cálculo puede refinarse teniendo en cuenta que $v(1) = e^1 \approx 2.718$ m/s, luego podemos decir que en el primer segundo el coche habrá recorrido aproximadamente 1 m y en el segundo segundo 2.718 m, luego $x(2) \approx x_0 + 3.718$ m.

Si aplicamos sistemáticamente esta idea, podemos obtener una aproximación mejor teniendo en cuenta la velocidad del coche cada décima de segundo: consideramos todos los instantes $t_i = 0.1i$, para $i = 0, \dots, 20$ y aproximamos el espacio recorrido en el intervalo $[t_{i-1}, t_i]$ por $v(t_{i-1}) \cdot \Delta t$, donde $\Delta t = 0.1$ s.

El resultado es

$$x(2) - x(0) \approx \sum_{i=1}^{20} v(t_{i-1})\Delta t = \sum_{i=1}^{20} e^{0.01i^2} 0.1 \approx 13.95 \text{ m.}$$

Vemos que el espacio recorrido es sustancialmente mayor que los 2 m que habíamos estimado inicialmente.

Antes de seguir, conviene plantear el problema en general:

Supongamos que una función $f : [a, b] \rightarrow \mathbb{R}$ es la derivada¹ de otra función $F : [a, b] \rightarrow \mathbb{R}$ y queremos calcular $F(b)$ a partir de f y de $F(a)$ o, equivalentemente, queremos calcular $F(b) - F(a)$ a partir de f .

Lo que hemos razonado es que una aproximación de $F(b) - F(a)$ puede obtenerse dividiendo $[a, b]$ en N partes iguales de longitud $\Delta x = (b - a)/N$ mediante los puntos $x_i = a + i(b - a)/N$, para $i = 0, \dots, N$, y considerando que

$$F(b) - F(a) \approx \sum_{i=1}^N f(x_{i-1})\Delta x.$$

Volviendo al ejemplo del coche, la tabla siguiente muestra la aproximación que obtenemos para distintos valores de N correspondientes a décimas de segundo, centésimas, milésimas, etc.:

N	20	200	2 000	20 000	200 000	2 000 000
$x(2) - x(0)$	13.9541	16.1865	16.4258	16.4499	16.4524	16.4526

Puede probarse que el límite cuando N tiende a ∞ es

$$x(2) - x(0) = 16.4526277655072302247364 \dots$$

con lo que el coche ha recorrido unos 16.5 metros en 2 segundos. La velocidad inicial era de 1 m/s y al final se movía a $e^4 \approx 54.6$ m/s (196.5 km/h), mientras que su velocidad media ha sido de 59.4 km/h.

En realidad, aun aceptando que la sucesión converja, no tenemos ninguna garantía de que el límite vaya a ser precisamente $x(2) - x(0)$, aunque sea razonable conjeturarlo. Pronto demostraremos que es así, pero de momento nos interesa más tener una idea clara de lo que estamos calculando. Lo que estamos planteando es que si tenemos una función $f(x)$ y queremos calcular cualquier valor $F(b)$ de una función F tal que $F' = f$ conociendo otro valor $F(a)$, podemos dividir el intervalo $[a, b]$ en un número grande de intervalos de longitud pequeña Δx , aproximar la variación de F en cada intervalo $[x_{i-1}, x_i]$ por la expresión $f(x)\Delta x$ y sumar dichas variaciones. El resultado será una aproximación a $F(b) - F(a)$. No será exacta porque $f(x)\Delta x$ no es la variación exacta de F en $[x_{i-1}, x_i]$, pero se aproxima más cuanto menor es Δx . Por ello “cabe esperar” que cuanto menor sea Δx (o, equivalentemente, cuanto mayor sea el número de partes N), más se parecerá la suma al resultado correcto.

¹De momento podemos entender que ambas funciones están definidas en un intervalo abierto mayor y que es en éste donde una es la derivada de la otra.

Leibniz diría que para obtener el valor exacto de $F(b) - F(a)$ hay que considerar una partición de $[a, b]$ en infinitos intervalos de longitud infinitesimal dx , de modo que $dF = f(x) dx$ sea la variación infinitesimal de F con un error despreciable², de modo que si sumamos estos infinitos incrementos infinitesimales, cuando x recorre el intervalo $[a, b]$, obtendremos el valor exacto de $F(b) - F(a)$.

El párrafo anterior —incluida la nota al pie— no tiene sentido, pero explica por qué Leibniz usó la notación

$$\int_a^b f(x) dx$$

para representar el valor exacto $F(b) - F(a)$ que queremos calcular. Esta notación sugiere que “debería” poder obtenerse como suma (el primer signo es una S deformada, correspondiente a la inicial de la palabra latina *summa*) de los incrementos infinitesimales $f(x) dx$ que se obtienen al partir $[a, b]$ en infinitos intervalos de longitud infinitesimal dx .

A continuación introducimos la notación de Leibniz despojándola de toda referencia a cantidades infinitesimales.

2.1 Integrales definidas e indefinidas

Definición 2.1 Diremos que una función $F : I \rightarrow \mathbb{R}$ definida en un intervalo abierto I es una *primitiva* de una función $f : I \rightarrow \mathbb{R}$ si F es derivable en I y $F' = f$.

En estos términos, lo que afirma el teorema 1.16 es que si una función f admite una primitiva F en un intervalo I , entonces el conjunto de todas las primitivas de f en I está formado por las funciones de la forma $F + c$, para cada $c \in \mathbb{R}$. Llamaremos *integral indefinida* de la función f , y la representaremos por

$$\int f(x) dx$$

al conjunto de todas las primitivas de f en un intervalo prefijado. Si F es una de ellas, es habitual escribir

$$\int f(x) dx = F(x) + c.$$

Por ejemplo, es evidente que

$$\int 3x^2 dx = x^3 + c.$$

Conviene definir como sigue el concepto de primitiva en un intervalo cerrado:

²Por ejemplo, si $F(x) = x^3$ y $f(x) dx = 3x^2 dx$, el incremento infinitesimal exacto de F sería $3x^2 dx + 3x dx^2 + dx^3$, pero Leibniz diría que los términos $3x dx^2 + dx^3$ son despreciables porque siguen siendo infinitamente pequeños cuando se dividen entre dx , es decir, que son “infinitésimos de orden superior a dx ”.

Definición 2.2 Una función $F : [a, b] \rightarrow \mathbb{R}$ es una *primitiva* de otra función $f :]a, b[\rightarrow \mathbb{R}$ si es continua en $[a, b]$, derivable en $]a, b[$ y $F' = f$ en el intervalo abierto. En tal caso, definimos la *integral definida* de f en $[a, b]$ como

$$\int_a^b f(x) dx = [F(x)]_a^b = F(b) - F(a).$$

Notemos que este valor es independiente de la primitiva con la que lo calculamos, pues si F_1 y F_2 cumplen lo mismo, el teorema 1.16 nos asegura que $F_2(x) = F_1(x) + c$ para cierta constante c y todo $x \in]a, b[$, pero tomando límites cuando x tiende a a y a b concluimos que esta relación se cumple también en a y en b , luego $F_2(b) - F_2(a) = F_1(b) - F_1(a)$.

La notación $[F(x)]_a^b = F(b) - F(a)$ es útil para expresar el paso intermedio en el cálculo de una integral definida correspondiente al momento en que hemos calculado la primitiva, pero todavía no hemos sustituido en ella los extremos de la integral.

Por ejemplo:

$$\int_1^4 3x^2 dx = [x^3]_1^4 = 4^3 - 1^3 = 63.$$

Es conveniente considerar que la definición de integral definida vale igualmente si $b \leq a$, de modo que, por definición,

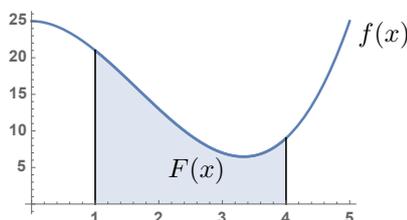
$$\int_b^a f(x) dx = F[a] - F[b] = - \int_a^b f(x) dx.$$

Nota Hay que insistir en que, técnicamente, la notación $\int_a^b () dx$ es “un bloque indivisible” como lo es la notación $d()/dx$ que hemos introducido para las derivadas. No obstante, la notación sugiere que una integral “debería” poder calcularse dividiendo el intervalo $[a, b]$, no ya en infinitos intervalos infinitesimales, pero sí en un número grande de intervalos de longitud pequeña Δx y sumando las cantidades $f(x_{i-1}) \Delta x$, de modo que podemos obtener aproximaciones arbitrariamente buenas de la integral tomando Δx suficientemente pequeño.

En realidad esto no es así en general. Por una parte, hay funciones f que no tienen primitiva, y hay funciones con primitiva para las que las “aproximaciones” que estamos describiendo no convergen a la primitiva. Sin embargo, vamos a demostrar que si f es continua todo funciona según lo esperado, es decir, que f tiene primitiva y que se puede calcular como límite de las aproximaciones que estamos considerando. ■

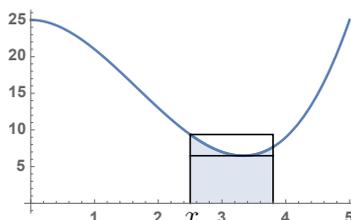
En primer lugar consideramos funciones que no tomen valores negativos y vamos a probar una sorprendente conexión entre el concepto de primitiva y el cálculo de áreas:

Teorema 2.3 Sea $f : [a, b] \rightarrow [0, +\infty[$ una función continua y $F : [a, b] \rightarrow \mathbb{R}$ la función que cumple que $F(x)$ es el área³ de la figura comprendida entre la gráfica de f , el eje $y = 0$ y las rectas horizontales de abscisas a y x . Entonces F es una primitiva de f en $[a, b]$.



DEMOSTRACIÓN: Por el teorema de Heine-Cantor [ITAn A.14], dado $\epsilon > 0$, existe un $\delta > 0$ tal que si $x, x' \in [a, b]$ cumplen $|x - x'| < \delta$, entonces se cumple también que $|f(x) - f(x')| < \epsilon$.

Sea $x \in]a, b[$ y sea $h \neq 0$ tal que $x + h \in]a, b[$ y $|h| < \delta$. Consideramos el intervalo $[x, x + h]$ (entendiendo que se trata del intervalo $[x + h, x]$ si $h < 0$). Por el teorema de Weierstrass [ITAn 3.31] existen puntos $x_1, x_2 \in [x, x + h]$ tales que, para todo $\xi \in [x, x + h]$, se cumple $f(x_1) \leq f(\xi) \leq f(x_2)$.



Si $h > 0$, entonces $|F(x + h) - F(x)|$ es el área de la figura limitada por la gráfica de f , el eje $y = 0$ y las rectas verticales de abscisas x y $x + h$. Dicha figura contiene al rectángulo de base $|h|$ y altura $f(x_1)$ y a su vez está contenida en el rectángulo de base $|h|$ y altura $f(x_2)$. Por lo tanto,

$$f(x_1)|h| \leq |F(x + h) - F(x)| \leq f(x_2)|h|.$$

Esto implica que

$$f(x_1) \leq \frac{|F(x + h) - F(x)|}{|h|} \leq f(x_2),$$

pero de aquí podemos pasar a

$$f(x_1) \leq \frac{F(x + h) - F(x)}{h} \leq f(x_2),$$

pues el cociente es siempre positivo.

³La demostración no requiere más que propiedades intuitivamente obvias del concepto de área, pero éste admite una definición precisa en términos de la medida de Jordan definida en el apéndice A de [ITAn]. Concretamente, la figura considerada en el enunciado es medible Jordan (es decir, tiene un área definida) por los teoremas [ITAn A.13], [ITAn A.15].

A su vez:

$$-(f(x) - f(x_1)) \leq \frac{F(x+h) - F(x)}{h} - f(x) \leq f(x_2) - f(x).$$

Como $x, x_1, x_2 \in [x, x+h]$ y $|h| < \delta$, se cumple que $|x - x_1|, |x - x_2| < \delta$, luego $|f(x) - f(x_1)|, |f(x_2) - f(x)| < \epsilon$, aunque en realidad podemos quitar los valores absolutos, y así

$$-\epsilon < \frac{F(x+h) - F(x)}{h} - f(x) < \epsilon,$$

o equivalentemente,

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| < \epsilon.$$

Esto prueba que existe

$$F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = f(x),$$

para todo punto $x \in]a, b[$. En particular esto implica que F es continua en $]a, b[$, pero falta probar que también es continua en a y en b . Veamos que es continua en b . El caso de a es similar. Para ello consideramos el valor máximo M que la función f toma en el intervalo $[a, b]$, que existe por el teorema de Weierstrass. Dado $\epsilon > 0$, si $x \in [a, b]$ cumple $|x - b| < \epsilon/M$, entonces $|F(b) - F(x)|$ es $0 < \epsilon$ si $x = b$ o bien es el área de la figura comprendida entre la gráfica de f , el eje $y = 0$ y las rectas verticales de abscisas x y b , la cual está contenida en el rectángulo de base $b - x$ y altura M , luego

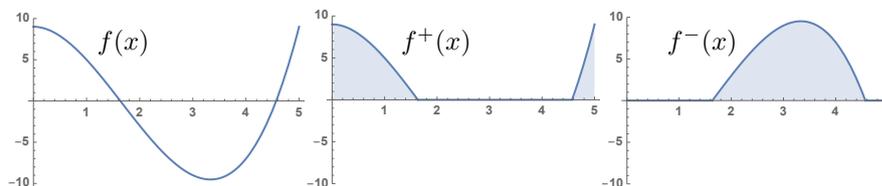
$$|F(b) - F(x)| \leq M(b - x) < \epsilon.$$

Esto prueba la continuidad en b . ■

Este resultado se conoce como *teorema fundamental del cálculo*, porque relaciona el concepto de integral como área con el concepto de integral como primitiva. Para el caso de funciones continuas no necesariamente positivas observamos que si $f : I \rightarrow \mathbb{R}$ es cualquier función continua, donde $I \subset \mathbb{R}$, también son continuas las funciones

$$f^+(x) = \max\{f(x), 0\}, \quad f^-(x) = -\min\{f(x), 0\}$$

y $f = f^+ - f^-$.



En efecto, basta tener en cuenta que

$$f^+(x) = \frac{|f(x)| + f(x)}{2}, \quad f^-(x) = \frac{|f(x)| - f(x)}{2}.$$

Con esto es fácil probar:

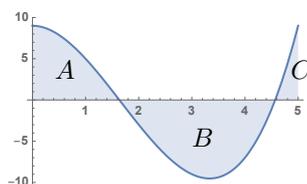
Teorema 2.4 *Toda función continua $f : [a, b] \rightarrow \mathbb{R}$ en un intervalo cerrado tiene una primitiva $F : [a, b] \rightarrow \mathbb{R}$, y la integral definida $\int_a^b f(x) dx$ se interpreta como la diferencia entre el área limitada por la gráfica de f^+ y el eje $y = 0$ menos el área limitada por la gráfica de f^- y el eje $y = 0$.*

DEMOSTRACIÓN: Basta descomponer $f = f^+ - f^-$, de modo que las funciones f^+ y f^- no toman valores negativos, luego por el teorema anterior tienen primitivas F_+ y F_- , de modo que $F'_+(x) = f^+(x)$ y $F'_-(x) = f^-(x)$, luego $F = F_+ - F_-$ cumple $F' = f^+ - f^- = f$. Además, según el teorema anterior, $F_+(b) - F_+(a) = F_+(b)$ es el área limitada entre la gráfica de f^+ y el eje $y = 0$, e igualmente con F_- , luego

$$\int_a^b f(x) dx = F_+(b) - F_-(b) - (F_+(a) - F_-(a)) = F_+(b) - F_-(b)$$

tiene la interpretación descrita en el enunciado. ■

Por ejemplo, la integral de la función que muestra la gráfica siguiente en el intervalo $[0, 5]$ es la suma de las áreas A y C menos el área B :



Nota Si $f : I \rightarrow \mathbb{R}$ es una función continua en un intervalo abierto I , es fácil ver que tiene primitiva $F : I \rightarrow \mathbb{R}$. En efecto, basta tomar un punto $a \in I$ y considerar la función

$$F(x) = \int_a^x f(t) dt.$$

En efecto, si $x \in I$ cumple $x > a$, podemos tomar $b \in I$ tal que $a < x < b$ y tenemos probado que F es una primitiva de f en el intervalo $[a, b]$, luego en particular $F'(x) = f(x)$. Una leve adaptación de la prueba muestra que lo mismo vale si $x < a$, luego $F' = f$ en todo I salvo a lo sumo en el punto a , pero, teniendo en cuenta que f es continua, el teorema 1.29 implica que F también es derivable en a y que $F' = f$ en todo el intervalo I . ■

Finalmente vamos a probar que la integral de una función continua se puede aproximar con la técnica que hemos discutido en la introducción a este capítulo. De hecho, probamos algo bastante más general:

Teorema 2.5 Si $f : [a, b] \rightarrow \mathbb{R}$ es una función continua, para cada $\epsilon > 0$ existe un $\delta > 0$ tal que si

$$a = x_0 < x_1 < \cdots < x_n < b$$

es cualquier partición de $[a, b]$ en intervalos tales que $\Delta x_i = x_i - x_{i-1} < \delta$ y $x_{i-1} \leq \xi_i \leq x_i$, entonces

$$\left| \sum_{i=1}^n f(\xi_i) \Delta x_i - \int_a^b f(x) dx \right| < \epsilon.$$

DEMOSTRACIÓN: Sea $F : [a, b] \rightarrow \mathbb{R}$ una primitiva de f . Entonces

$$\int_a^b f(x) dx = F(b) - F(a) = \sum_{i=1}^n (F(x_i) - F(x_{i-1})).$$

Además, por el teorema del valor medio, existe $x_{i-1} < \eta_i < x_i$ tal que

$$F(x_i) - F(x_{i-1}) = f(\eta_i) \Delta x_i.$$

Por consiguiente,

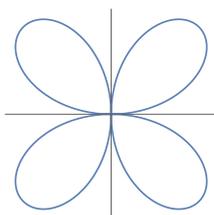
$$\begin{aligned} \left| \sum_{i=1}^n f(\xi_i) \Delta x_i - \int_a^b f(x) dx \right| &= \left| \sum_{i=1}^n f(\xi_i) \Delta x_i - \sum_{i=1}^n f(\eta_i) \Delta x_i \right| \\ &\leq \sum_{i=1}^n |f(\xi_i) - f(\eta_i)| \Delta x_i. \end{aligned}$$

Ahora usamos el teorema de Heine-Cantor [ITAn A.14], según el cual, dado $\epsilon > 0$, existe un $\delta > 0$ tal que si $x, y \in [a, b]$ cumplen $|x - y| < \delta$, entonces $|f(x) - f(y)| < \epsilon/(b - a)$. Si se cumple que $\Delta x_i < \delta$ para todo i , entonces, como $x_{i-1} \leq \xi_i, \eta_i \leq x_i$, también $|\xi_i - \eta_i| < \delta$, luego $|f(\xi_i) - f(\eta_i)| < \epsilon/(b - a)$, y podemos concluir que

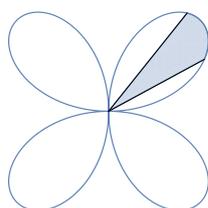
$$\left| \sum_{i=1}^n f(\xi_i) \Delta x_i - \int_a^b f(x) dx \right| < \frac{\epsilon}{b - a} \sum_{i=1}^n \Delta x_i = \frac{\epsilon}{b - a} (b - a) = \epsilon. \quad \blacksquare$$

Con esto ya tenemos demostrado que la solución del problema planteado en la introducción de este capítulo es el valor $x(2) - x(0) = 16.4526\dots$ que puede obtenerse dividiendo el intervalo $[0, 2]$ en intervalos pequeños y aproximando el espacio recorrido en cada uno de ellos mediante la velocidad del coche en el instante inicial del intervalo (de hecho, hemos probado que daría igual considerar la velocidad del coche en cualquier otro punto del intervalo).

Veamos una aplicación del teorema anterior. Cualquier función continua $\rho : [0, 2\pi] \rightarrow [0, +\infty[$ que cumpla $\rho(0) = \rho(2\pi)$ determina una curva en coordenadas polares que rodea al punto $(0, 0)$. Por ejemplo, $\rho = \text{sen } 2\theta$ determina una curva en forma de flor:



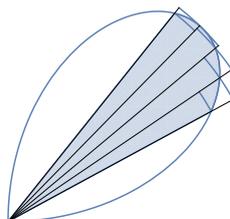
Vamos a calcular el área limitada por una curva en coordenadas polares. Más en general, vamos a determinar el área limitada entre dos radios cualesquiera de coordenadas $\theta_0 < \theta_1$:



Para ello dividimos el intervalo $[\theta_0, \theta_1]$ en subintervalos:

$$\theta_0 = t_0 < t_1 < \dots < t_n = \theta_1.$$

Llamamos $\Delta t_i = t_i - t_{i-1}$ y, en virtud del teorema de Weierstrass, existen valores θ_i^0, θ_i^1 en el intervalo $[t_{i-1}, t_i]$ tales que $r_i = \rho(\theta_i^0)$ y $R_i = \rho(\theta_i^1)$ son, respectivamente, el valor mínimo y el valor máximo de la función $\rho(\theta)$ en el intervalo $[t_{i-1}, t_i]$. Es claro entonces que el fragmento de la figura cuya área queremos calcular formada por los puntos con $t_{i-1} \leq \theta \leq t_i$ contiene el sector circular de radio r_i y amplitud $\Delta t_i = t_i - t_{i-1}$ y está contenido en el sector circular de radio R_i y amplitud Δt_i :



Por consiguiente, el área A_i de dicho fragmento cumple

$$\pi r_i^2 \frac{\Delta t_i}{2\pi} \leq A_i \leq \pi R_i^2 \frac{\Delta t_i}{2\pi},$$

luego el área A que queremos calcular cumple

$$\frac{1}{2} \sum_{i=1}^n \rho^2(\theta_i^0) \Delta t_i \leq A \leq \frac{1}{2} \sum_{i=1}^n \rho^2(\theta_i^1) \Delta t_i,$$

y el teorema anterior nos asegura que, para todo $\epsilon > 0$, existe un $\delta > 0$ tal que si $\Delta t_i < \delta$ para todo i , entonces

$$\frac{1}{2} \int_{\theta_0}^{\theta_1} \rho^2(\theta) d\theta - \epsilon \leq \frac{1}{2} \sum_{i=1}^n \rho^2(\theta_i^0) \Delta t_i \leq A \leq \frac{1}{2} \sum_{i=1}^n \rho^2(\theta_i^1) \Delta t_i \leq \frac{1}{2} \int_{\theta_0}^{\theta_1} \rho^2(\theta) d\theta + \epsilon,$$

luego

$$\left| A - \frac{1}{2} \int_{\theta_0}^{\theta_1} \rho^2(\theta) d\theta \right| \leq \epsilon.$$

Así hemos probado el teorema siguiente:

Teorema 2.6 *El área limitada por la curva formada por los puntos cuyas coordenadas polares son de la forma $(\rho, \rho(\theta))$ y por las semirrectas determinadas por $\theta_0 < \theta_1$ viene dada por*

$$A = \frac{1}{2} \int_{\theta_0}^{\theta_1} \rho^2(\theta) d\theta.$$

Ejemplo El área de un “pétalo” de la curva dada por $\rho = \text{sen } 2\theta$ se obtiene integrando ρ^2 entre 0 y $\pi/2$, es decir:

$$A = \frac{1}{2} \int_0^{\pi/2} \text{sen}^2 2\theta d\theta.$$

Para calcular esta integral observamos que

$$\text{sen}^2 \theta = \frac{1 - \cos 4\theta}{2},$$

y que una primitiva de esta función es claramente

$$\frac{\theta}{2} - \frac{\text{sen } 4\theta}{8},$$

luego

$$A = \frac{1}{2} \left[\frac{\theta}{2} - \frac{\text{sen } 4\theta}{8} \right]_0^{\pi/2} = \frac{\pi}{8}.$$

y la “flor” completa tiene área $\pi/2$. ■

Nota El teorema anterior aplicado a $\rho(\theta) = r$ nos da que el área de un círculo de radio r es πr^2 , pero en realidad con ello no hemos probado nada, porque en la prueba del teorema hemos usado que el área de un sector circular de amplitud θ es $\pi r^2 \theta / 2\pi = r^2 \theta / 2$, luego estamos suponiendo conocido el valor del área del círculo. En la sección siguiente veremos cómo calcular el área de un círculo mediante una integral. ■

El lector tiene otra aplicación del teorema anterior en la sección B.5, donde calculamos el área de la lemniscata de Bernoulli.

Terminamos esta sección con algunas propiedades básicas de las integrales definidas:

Teorema 2.7 Si f y g son funciones con primitiva en los intervalos en los que se integra en el miembro derecho de cada una de las fórmulas de los apartados a)–d), entonces el integrando del miembro izquierdo también tiene primitiva y se cumple la fórmula:

$$a) \int_a^b (\alpha f(x) + \beta g(x)) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx.$$

$$b) \int_a^b f(x) dx = - \int_b^a f(x) dx.$$

$$c) \int_a^a f(x) dx = 0.$$

$$d) \int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx.$$

$$e) \text{ Si } f(x) \leq g(x) \text{ en } [a, b] \text{ y ambas tienen primitiva, } \int_a^b f(x) dx \leq \int_a^b g(x) dx.$$

$$f) \text{ Si } f \text{ y } |f| \text{ tienen primitiva en } [a, b], \text{ entonces } \left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx.$$

DEMOSTRACIÓN: La propiedad a) es inmediata, pues afirma esencialmente que si F y G son primitivas de f y g , respectivamente, entonces $\alpha F + \beta G$ es una primitiva de $\alpha f + \beta g$. La propiedad b) es una mera consecuencia de que hemos definido

$$\int_a^b f(x) dx = F(b) - F(a)$$

sin exigir que sea $a < b$, con lo que si intercambiamos los extremos de la integral ésta cambia de signo. Igualmente, la propiedad c) es una consecuencia de la definición.

En la propiedad d) hay varias sutilezas que tener en cuenta. Una es que, en virtud de b), la relación es válida aunque no se cumpla $a < c < b$, pero todos los casos posibles se reducen a éste cambiando integrales de miembro si es necesario. La otra es que, en este caso en concreto, si F_1 es una primitiva de f en $[a, b]$ y F_2 es una primitiva de f en $[b, c]$, sumándole a F_2 una constante podemos suponer que $F_1(b) = F_2(b)$, con lo que ambas funciones se extienden a una función continua $F : [a, b] \rightarrow \mathbb{R}$ que cumple $F'(x) = f(x)$ salvo a lo sumo en $x = c$, pero entonces el teorema 1.29 implica que F también es derivable en b y $F'(b) = f(b)$.

Para probar e) observamos que si F y G son primitivas de f y g , respectivamente, entonces $(G - F)'(x) = g(x) - f(x) \geq 0$, luego la función $G - F$ es monótona creciente⁴ en $[a, b]$, luego $G(a) - F(a) \leq G(b) - F(b)$, luego concluimos que $F(b) - F(a) \leq G(b) - G(a)$.

⁴En principio es monótona creciente en $]a, b[$, pero, por el teorema del valor medio, si $x \in]a, b[$, existe $a < c < x$ tal que $f(x) - f(a) = f'(c)(x - a) \geq 0$, e igualmente se razona con b , por lo que f es monótona creciente en $[a, b]$.

La propiedad f) se deduce de e), pues $-|f(x)| \leq f(x) \leq |f(x)|$, luego

$$-\int |f(x)| dx \leq \int f(x) dx \leq \int |f(x)| dx,$$

de donde se sigue la relación del enunciado. \blacksquare

El teorema siguiente afirma que, si se cumplen ciertas hipótesis, podemos intercambiar un límite con una integral:

$$\int_a^b \lim_n f_n(x) dx = \lim_n \int_a^b f_n(x) dx.$$

Teorema 2.8 Sea $f_n : [a, b] \rightarrow \mathbb{R}$ una sucesión de funciones continuas y sea $f : [a, b] \rightarrow \mathbb{R}$ otra función, también continua, de modo que, para todo $x \in [a, b]$ y todos los números naturales $m \leq n$, se cumpla que $f_m(x) \leq f_n(x) \leq f(x)$ y $\lim_n f_n(x) = f(x)$. Entonces

$$\int_a^b f(x) dx = \lim_n \int_a^b f_n(x) dx.$$

DEMOSTRACIÓN: Basta probar que la sucesión

$$\int_a^b (f(x) - f_n(x)) dx = \int_a^b f(x) dx - \int_a^b f_n(x) dx$$

converge a 0. Para ello llamamos $g_n(x) = f(x) - f_n(x)$, de modo que si $x \in [a, b]$ y $m \leq n$ se cumple $g_m(x) \geq g_n(x) \geq 0$ y $\lim_n g_n(x) = 0$. Tenemos que probar que la sucesión de las integrales converge a 0.

Para ello, dado $\epsilon > 0$, basta probar que existe un k tal que si $n \geq k$, se cumple que $g_n(x) \leq \epsilon/(b-a)$ para todo $x \in [a, b]$, pues, aceptando esto, tenemos que

$$0 \leq \int_a^b g_n(x) dx \leq \int_a^b \frac{\epsilon}{b-a} dx = \epsilon.$$

Supongamos, por el contrario, que para todo k existe un $n_k \geq k$ y un punto $x_k \in [a, b]$ de modo que $g_{n_k}(x_k) > \eta$ (donde $\eta = \epsilon/(b-a)$). Claramente, podemos construir una sucesión $n_0 < n_1 < n_2 < \dots$ de números naturales y una sucesión de puntos x_k de modo que $g_{n_k}(x_k) > \eta$.

Por el teorema de Bolzano-Weierstrass [ITAn 1.24] pasando a una subsucesión, podemos suponer que existe $\lim_k x_k = x^* \in [a, b]$. Como $\lim_n g_n(x^*) = 0$, existe un n tal que $g_n(x^*) < \eta/4$. Como g_n es continua, existe un $\delta > 0$ tal que si $|x - x^*| < \delta$, entonces $|g_n(x) - g_n(x^*)| < \eta/4$, con lo que

$$g_n(x) \leq |g_n(x) - g_n(x^*)| + g_n(x^*) < \frac{\eta}{4} + \frac{\eta}{4} = \frac{\eta}{2}.$$

Sea $k \geq n$ tal que $|x_{n_k} - x^*| < \delta$, con lo que $g_{n_k}(x_{n_k}) \leq g_n(x_{n_k}) < \eta/2$, y tenemos una contradicción. \blacksquare

2.2 Cálculo de primitivas

Mientras la derivación de funciones “usuales” se reduce a aplicar sistemáticamente las reglas de derivación, no existe ningún procedimiento para obtener una primitiva de una función “usual” dada. De hecho, hay funciones muy sencillas, como e^{x^2} para las que puede probarse que, aunque tienen primitivas, por ser continuas, éstas no son expresables en términos de funciones “usuales”. Existen diversas técnicas para integrar determinadas clases de funciones, pero incluso en estos casos no es raro que un razonamiento *ad hoc* permita obtener el resultado mucho más fácilmente que aplicando un procedimiento general.

Dado que hay mucha bibliografía dedicada específicamente al cálculo de primitivas, no intentaremos aquí ayudar al lector a adquirir cierta competencia en tal disciplina, sino que en esta sección nos limitaremos a discutir los aspectos más relevantes desde un punto de vista conceptual y en el apéndice C describiremos brevemente las técnicas de integración y recopilaremos allí todas las primitivas que necesitaremos a lo largo de este libro. Así, el lector que no esté interesado en el cálculo de primitivas puede seguir los argumentos libres de cálculos intercalados que, en cualquier caso, puede consultar —si lo desea— en el apéndice, y el lector interesado en el cálculo de primitivas encontrará todos los ejemplos reunidos y clasificados para poder compararlos y asimilarlos mejor.

Integrales inmediatas Cada regla de derivación da lugar a una regla de integración. Por ejemplo, el hecho de que la derivada de a^x sea $a^x \log a$, se traduce en que una primitiva de a^x es $a^x / \log a$ o, equivalentemente:

$$\int a^x dx = \frac{1}{\log a} a^x + c.$$

Esto es una regla de integración, pero para aumentar su alcance conviene combinarla con la regla de la cadena y observar que, si $f(x)$ es una función arbitraria, entonces la derivada de $a^{f(x)}$ es $a^{f(x)} f'(x) \log a$, lo que nos da la regla más general:

$$\int f'(x) a^{f(x)} dx = \frac{1}{\log a} a^{f(x)} + c.$$

Ésta es una de las reglas que figuran en la tabla C.1, y del mismo modo que hemos justificado ésta se justifican todas las demás. Basta derivar su miembro derecho con las reglas de derivación usuales y comprobar que el resultado es el integrando correspondiente.

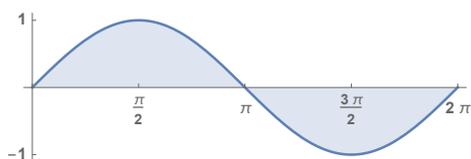
Remitimos al lector al principio de la sección C.1 para una breve discusión sobre los aspectos más elementales que conviene conocer sobre el cálculo de integrales inmediatas, y a continuación mostramos algunos ejemplos menos evidentes de manipulaciones que permiten reducir algunas integrales a las reglas de la tabla.

Naturalmente, el cálculo de integrales indefinidas nos permite a su vez calcular integrales definidas, que a su vez pueden interpretarse como cálculo de

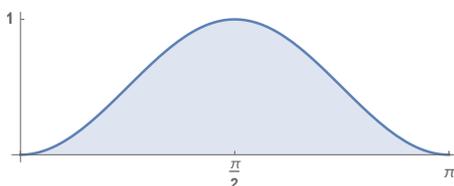
áreas. Por ejemplo, ahora es evidente que

$$\int_0^{\pi} \operatorname{sen} x \, dx = [-\cos x]_0^{\pi} = 2,$$

lo que significa que el área encerrada por la gráfica del seno entre 0 y π mide 2 unidades cuadradas:



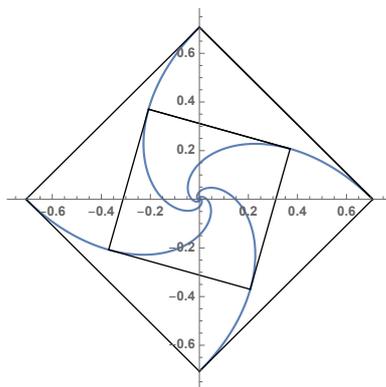
En cambio, no es tan inmediato cómo calcular el área que limitan las funciones $\operatorname{sen}^2 x$ y $\operatorname{cos}^2 x$:



La respuesta es que el área es $\pi/2$, y el cálculo está en (C.2).

Ejemplo: Persecución cíclica Sea $n \geq 3$ y supongamos que n ratones se encuentran en los vértices de un polígono regular de n lados y que, en un mismo instante, todos empiezan a caminar a una misma velocidad constante dirigiéndose siempre hacia el punto en el que se encuentra el ratón siguiente. Determinar la trayectoria que seguirán los ratones y el tiempo que tardarán en encontrarse.

Supondremos que el lado del polígono mide 1 unidad y que los ratones se mueven a velocidad unitaria. (La solución general se obtiene fácilmente a partir de la de este caso.)



La figura muestra las trayectorias para $n = 4$, así como la situación de los ratones en un instante intermedio. Observemos que, dado que la situación de los ratones es simétrica, en cada instante todos ellos estarán a la misma distancia del centro del cuadrado inicial, y todos habrán girado un mismo ángulo alrededor de él, lo cual se traduce en que en todo momento estarán situados sobre los vértices de un polígono regular de n lados, con centro en el centro del polígono inicial, pero con radio cada vez menor.

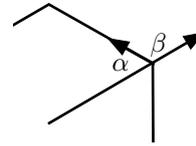
Consideremos un sistema de referencia con origen en el centro del polígono y expresemos la posición de uno de los ratones en coordenadas polares:

$$r = (x, y) = \rho(\cos \theta, \operatorname{sen} \theta).$$

Entonces la velocidad es

$$v = \rho'(\cos \theta, \operatorname{sen} \theta) + \rho\theta'(-\operatorname{sen} \theta, \cos \theta).$$

El hecho de que los ratones estén en los vértices de un polígono regular de n lados se traduce en que la velocidad de cualquiera de ellos se sitúa sobre uno de los lados y, por consiguiente, forma un ángulo β con el vector unitario $(\cos \theta, \operatorname{sen} \theta)$ tal que $\alpha = \pi - \beta = \pi/2 - \pi/n$, luego



$$\rho' = (\cos \theta, \operatorname{sen} \theta) \cdot v = \cos \beta = -\cos \alpha = -\operatorname{sen} \frac{\pi}{n}.$$

El radio de un polígono de lado unitario es $\rho(0) = (2 \operatorname{sen}(\pi/n))^{-1}$, luego

$$\rho(t) = \frac{1}{2 \operatorname{sen}(\pi/n)} - t \operatorname{sen}(\pi/n) = \frac{1 - 2t \operatorname{sen}^2(\pi/n)}{2 \operatorname{sen}(\pi/n)}.$$

Esto nos da ya el tiempo que tardarán los ratones en encontrarse:

$$T = \frac{1}{2 \operatorname{sen}^2(\pi/n)} = \frac{1}{1 - \cos(2\pi/n)},$$

que es también la distancia que recorren, ya que se mueven a velocidad unitaria. Ahora imponemos que el módulo de la velocidad tiene que ser 1:

$$1 = v \cdot v = \rho'^2 + \rho^2 \theta'^2 = \operatorname{sen}^2(\pi/n) + \left(\frac{1 - 2t \operatorname{sen}^2(\pi/n)}{2 \operatorname{sen}(\pi/n)} \right)^2 \theta'^2.$$

Despejando queda

$$\theta' = \frac{2 \operatorname{sen}(\pi/n) \cos(\pi/n)}{1 - 2t \operatorname{sen}^2(\pi/n)},$$

luego

$$\begin{aligned} \theta(t) &= \theta_0 + 2 \operatorname{sen}(\pi/n) \cos(\pi/n) \int_0^t \frac{du}{1 - 2u \operatorname{sen}^2(\pi/n)} = \\ &= \theta_0 - \frac{2 \operatorname{sen}(\pi/n) \cos(\pi/n)}{2 \operatorname{sen}^2(\pi/n)} \int_0^t \frac{-2 \operatorname{sen}^2(\pi/n)}{1 - 2u \operatorname{sen}^2(\pi/n)} du = \\ &= \theta_0 - \cot(\pi/n) \log(1 - 2t \operatorname{sen}^2(\pi/n)). \end{aligned}$$

Con esto ya tenemos determinada la trayectoria de los ratones. Por ejemplo, para $n = 4$ las fórmulas se simplifican hasta

$$\rho(t) = \frac{\sqrt{2}}{2}(1-t), \quad \theta(t) = \theta_0 - \log(1-t),$$

con lo que las trayectorias son

$$r(t) = \frac{\sqrt{2}}{2}(1-t)(\cos(\theta_0 - \log(1-t)), \sin(\theta_0 - \log(1-t))),$$

y la distancia que recorre cada ratón hasta encontrarse con los otros es igual a la longitud del lado. ■

Ejemplo Una cinta elástica tensa de 1 metro está sujeta por un extremo, y el extremo opuesto se mueve a una velocidad de 2 metros por minuto. En el extremo fijo se encuentra una hormiga que camina a una velocidad de 1 metro por minuto y se dirige hacia el otro extremo. ¿Lo alcanzará? y, en caso afirmativo, ¿en cuánto tiempo?

En general, si la cinta tiene una longitud inicial l_0 y su extremo se mueve a velocidad v_e , su longitud en cada instante es $l_0 + v_e t$. Sea $x(t)$ la distancia de la hormiga al extremo fijo y llamemos

$$r(t) = \frac{x(t)}{l_0 + v_e t}$$

a la fracción de cinta que ha recorrido la hormiga en cada instante. Si la hormiga estuviera en reposo, el valor de $r(t)$ sería constante a pesar del estiramiento de la cinta. Si la hormiga se mueve a una velocidad v_h esto significa que $r(t)$ aumenta a una velocidad

$$r'(t) = \frac{v_h}{l_0 + v_e t}.$$

Teniendo en cuenta que $r(0) = 0$, de aquí deducimos que

$$\begin{aligned} r(t) &= \int_0^t \frac{v_h}{l_0 + v_e t} dt = \frac{v_h}{v_e} \int_0^t \frac{v_e}{l_0 + v_e t} dt = \frac{v_h}{v_e} (\log(l_0 + v_e t) - \log l_0) \\ &= \frac{v_h}{v_e} \log \frac{l_0 + v_e t}{l_0}. \end{aligned}$$

Buscamos el instante T en que $r(T) = 1$, que claramente es

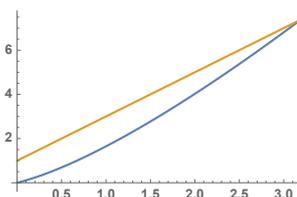
$$T = \frac{l_0}{v_e} (e^{v_e/v_h} - 1).$$

En este tiempo, la hormiga habrá recorrido una distancia

$$D = l_0 + v_e T = l_0 e^{v_e/v_h}.$$

sí pues, la hormiga siempre llega al extremo opuesto, sean cuales sean los datos del problema.

En las condiciones concretas del enunciado, con $v_e = 2$, $v_h = 1$, $l_0 = 1$, el tiempo requerido es $T = 0.5(e^2 - 1) = 3.2$ minutos y la distancia recorrida será $D = e^2 = 7.39$ metros. La gráfica muestra el movimiento de la hormiga y del extremo de la cinta.



Integración por partes La regla para el producto de derivadas afirma que si dos funciones $u(x)$ y $v(x)$ son derivables, entonces

$$u(x)v'(x) = (u(x)v(x))' - v(x)u'(x).$$

Como $u(x)v(x)$ es una primitiva de $(u(x)v(x))'$, si existe una primitiva de $v(x)u'(x)$, también existe una de $u(x)v'(x)$, que viene dada por:

$$\int u(x)v'(x) dx = u(x)v(x) - \int v(x)u'(x) dx.$$

Si convenimos en abreviar $dv = v'(x) dx$ y $du = u'(x) dx$, obtenemos la forma más habitual en la que se expresa la regla de integración por partes:

$$\int u dv = uv - \int v du.$$

No obstante, la idea subyacente se aprecia mejor en la expresión previa: si tenemos que integrar un producto $u(x)v'(x)$ y conocemos una primitiva $v(x)$ de uno de los factores, la regla reduce el problema a encontrar una primitiva de $v(x)u'(x)$, es decir, que podemos transformar la integral en otra en la que uno de los factores aparece derivado y el otro integrado.

De aquí se sigue la versión para integrales definidas de la integración por partes, que es:

$$\int_a^b u dv = [uv]_a^b - \int_a^b v du.$$

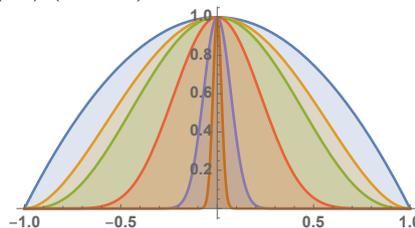
Remitimos al lector a la sección C.3 para ver ejemplos prácticos de aplicación de esta fórmula. Aquí vamos a ver un ejemplo más sofisticado:

Ejemplo Para cada número natural n , se cumple que

$$\int_{-1}^1 (1-x^2)^n dx = \frac{2^{2n+1}(n!)^2}{(2n)!(2n+1)}. \quad (2.1)$$

La figura muestra las gráficas de los integrandos para $n = 1, 2, 3, 10, 100, 1000$. Vemos que, cuanto mayor es n , más se estrecha la gráfica, por lo que la integral es cada vez menor. Para probarlo integramos por partes con

$$u = (1-x^2)^n, \quad dv = dx, \quad du = 2n(1-x^2)^{n-1} dx, \quad v = x,$$



con lo que obtenemos que

$$\int_{-1}^1 (1-x^2)^n dx = [(1-x^2)^n x]_{-1}^1 + 2n \int_{-1}^1 (1-x^2)^{n-1} x^2 dx$$

y el primer término se anula al evaluar x en -1 y 1 . Integrando de nuevo por partes con

$$u = (1-x^2)^{n-1}, \quad dv = x^2, \quad du = -2(n-1)(1-x^2)^{n-2} x dx, \quad v = \frac{x^3}{3},$$

obtenemos que

$$\int_{-1}^1 (1-x^2)^n dx = \frac{2^2 n(n-1)}{3} \int_{-1}^1 (1-x^2)^{n-2} x^4 dx$$

y, en general, una simple inducción prueba que

$$\int_{-1}^1 (1-x^2)^n dx = \frac{2^k n(n-1) \cdots (n-k+1)}{3 \cdot 5 \cdots (2k-1)} \int_{-1}^1 (1-x^2)^{n-k} x^{2k} dx.$$

En efecto, tenemos probada la fórmula para $k = 1, 2$ y, si es cierta para k , integrando de nuevo por partes se obtiene la fórmula para $k + 1$. En particular, para $k = n$ la fórmula se reduce a

$$\int_{-1}^1 (1-x^2)^n dx = \frac{2^n n!}{3 \cdot 5 \cdots (2n-1)} \int_{-1}^1 x^{2n} dx.$$

Ahora basta observar que el denominador es $(2n)!/2^n n!$ y que la última integral vale $2/(2n+1)$. ■

Integración por cambio de variable La regla de la cadena nos dice que si

$$\int f(x) dx = F(x) + c,$$

es decir, si $F'(x) = f(x)$, y tenemos una función derivable $x(t)$, entonces

$$F(x(t))' = F'(x(t))x'(t) = f(x(t))x'(t),$$

es decir, que

$$\int f(x(t))x'(t) dt = F(x(t)) + c.$$

Supongamos que la función $x(t)$ tiene inversa $t(x)$. Entonces, si queremos calcular $\int f(x) dx$, es decir, si queremos calcular $F(x)$, podemos tratar de calcular una integral de la forma

$$\int f(x(t))x'(t) dt = G(t) + c,$$

para cierta función $x(t)$, en cuyo caso habremos calculado $G(t) = F(x(t))$, luego podremos recuperar $F(x) = G(t(x))$.

Podemos expresar esto en la forma

$$\int f(x) dx = \int f(x(t))x'(t) dt$$

si entendemos que en el segundo miembro, que es una función de t , hay que cambiar t por $t(x)$ para tener el primer miembro. Más aún, si convenimos en abreviar $dx = x'(t) dt$, vemos que el paso del primer miembro al segundo se reduce a sustituir formalmente x por $x(t)$ y dx por $x'(t) dt$.

He aquí la versión para integrales definidas, en la que es esencial el convenio que hemos adoptado según el cual el signo de una integral depende del orden de los extremos de integración:

Teorema 2.9 (Teorema de cambio de variable) *Sea $f : [a, b] \rightarrow \mathbb{R}$ una función con primitiva en $[a, b]$ y sea $x : [c, d] \rightarrow [a, b]$ una función continua estrictamente creciente o decreciente y derivable en $]c, d[$, con derivada no nula (luego con inversa derivable $t(x)$). Entonces*

$$\int_a^b f(x) dx = \int_{t(a)}^{t(b)} f(x(t))x'(t) dt.$$

DEMOSTRACIÓN: Por la regla de la cadena, si $F(x)$ es una primitiva de $f(x)$ en $[a, b]$, entonces $F(x(t))$ es una primitiva de $f(x(t))x'(t)$ en $[c, d]$. Entonces

$$\int_a^b f(x) dx = F(b) - F(a) = F(x(t(b))) - F(x(t(a))) = \int_{t(a)}^{t(b)} f(x(t))x'(t) dt. \quad \blacksquare$$

Ejemplo Se cumple que

$$\int_{-r}^r \sqrt{r^2 - x^2} dx = \frac{\pi r^2}{2}.$$

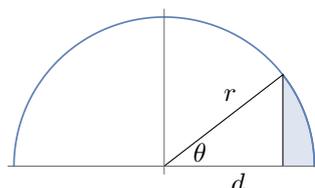
Esto es obvio a partir de la integral indefinida del ejemplo (C.14), pero si nos interesa la integral definida el cálculo puede simplificarse un poco. En efecto, aplicando el cambio de variable $x = r \cos t$ obtenemos:

$$\int_{-r}^r \sqrt{r^2 - x^2} dx = -r \int_{\pi}^0 \sqrt{r^2 - r^2 \cos^2 t} \sin t dt = r^2 \int_0^{\pi} \sin^2 t dt = \frac{\pi r^2}{2}.$$

De acuerdo con la interpretación geométrica de la integral, acabamos de calcular el área de un semicírculo de radio r , lo que nos permite concluir que el área de un círculo de radio r es πr^2 .

Los dos sumandos en los que hemos descompuesto la integral tienen una interpretación geométrica que se aprecia mejor si, fijado $0 < d = r \cos \theta < r$, calculamos

$$\begin{aligned} \int_d^r \sqrt{r^2 - x^2} dx &= -r \int_{\theta}^0 \sqrt{r^2 - r^2 \cos^2 t} \sin t dt = r^2 \int_0^{\theta} \sin^2 t dt \\ &= \frac{r^2 \theta}{2} - \frac{r^2}{4} \sin 2\theta = \frac{r^2 \theta}{2} - \frac{r^2}{2} \sin \theta \cos \theta = \frac{r^2 \theta}{2} - \frac{d}{2} \sqrt{r^2 - d^2}. \end{aligned}$$



El primer término es el área del sector circular de amplitud θ , y el segundo es el área del triángulo de base d y altura $r \operatorname{sen} \theta = \sqrt{r^2 - d^2}$. ■

Ejemplo: El área de una elipse Según [IGE 7.15], una elipse de semiejes a y b tiene ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

luego la función

$$y = \frac{b}{a} \sqrt{a^2 - x^2}$$

deja por debajo la mitad de la elipse, luego el área de la elipse completa es

$$A = 2 \int_{-a}^a \frac{b}{a} \sqrt{a^2 - x^2} dx = \frac{2b}{a} \frac{\pi a^2}{2} = \pi ab,$$

como ya calculamos al final de la sección 5.2 de [ITAn]. ■

Áreas de curvas paramétricas Vamos a ilustrar muchos de los resultados de este capítulo aplicándolos a cinco curvas: la cicloide, la cardioide, la nefroide, la tractriz y la lemniscata de Bernoulli. Para no diseminar los resultados, los hemos agrupado en el apéndice C. El lector puede, si lo desea, leer las definiciones de las cinco curvas y la forma en que se obtienen sus ecuaciones, salvo en el caso de la tractriz, que para llegar a ellas hay que resolver una ecuación diferencial. No obstante, basta tomar como definición que la tractriz es la curva dada por cualquiera de las parametrizaciones (B.3) o (B.4) para estar en condiciones de leer lo que sigue.

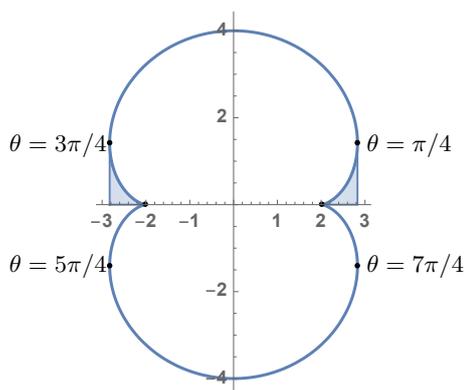
Además, con lo visto hasta aquí, el lector puede seguir ya el cálculo del área de la cardioide y la lemniscata, que se basan en el teorema 2.6, y también el cálculo de las áreas de la cicloide y la tractriz, mientras que para la nefroide conviene tener en cuenta la discusión siguiente:

En el cálculo del área de la cicloide y la tractriz, hemos empleado la misma técnica, a saber, hemos considerado la función $y(x)$ que expresa la coordenada y de un punto de la curva en términos de su coordenada x , y hemos expresado el área como

$$A = \int_a^b y(x) dx.$$

Esto no es más que la interpretación de la integral como área. Sin embargo, en ambos casos no conocíamos el integrando $y(x)$, pero eso no ha sido inconveniente

porque lo hemos eliminado mediante el cambio de variable $x = x(t)$, que deja la integral en términos de las funciones coordenadas de la parametrización conocida de la curva. Para que el cambio de variable sea correcto es necesario en principio que la función $x(t)$ sea creciente o decreciente, cosa que se cumple en el caso de las dos curvas, pero si intentamos hacer lo mismo con la nefroide vemos que ya no es el caso.



En efecto, el parámetro θ de la nefroide varía en $[0, 2\pi]$. Al recorrerla, partimos de $r(0) = (0, 2a)$, y mientras θ varía entre 0 y $\pi/4$ la función $x(\theta)$ es creciente, pero a partir de ahí decrece hasta $\theta = 3\pi/4$, luego crece hasta $\theta = \pi$, decrece hasta $\theta = 5\pi/4$, crece hasta $\theta = 7\pi/4$ y finalmente decrece hasta $\theta = 2\pi$. En particular, no existe una función $y(x)$ que nos de la coordenada y de un punto la nefroide con primera coordenada x , pues para una misma x puede haber dos o incluso cuatro puntos en la nefroide con esa primera coordenada. Sin embargo, vamos a ver que la misma integral

$$A = - \int_0^{2\pi} y(t)x'(t) dt \quad (2.2)$$

con la que hemos calculado el área de la cicloide y de la tractriz (sólo que cambiada de signo) nos da también el área de la nefroide. Para ello la descomponemos en los intervalos que hemos enumerado:

$$\begin{aligned} A = & - \int_0^{\pi/4} y(t)x'(t) dt - \int_{\pi/4}^{3\pi/4} y(t)x'(t) dt - \int_{3\pi/4}^{\pi} y(t)x'(t) dt \\ & - \int_{\pi}^{5\pi/4} y(t)x'(t) dt - \int_{5\pi/4}^{7\pi/4} y(t)x'(t) dt - \int_{7\pi/4}^{2\pi} y(t)x'(t) dt \end{aligned}$$

y vamos a interpretarlas geométricamente.

En el intervalo $[0, \pi/4]$, la nefroide es la gráfica de una función $y(x)$, y la función $x(t)$ es creciente, luego podemos usarla como cambio de variable para la integral

$$- \int_0^{x(\pi/4)} y(x) dx = - \int_0^{\pi/4} y(t)x'(t) dt,$$

que nos da el área sombreada en la parte derecha de la figura (que no forma parte del área abarcada por la nefroide). Con el signo negativo antepuesto, se convierte en el área en negativo.

Igualmente, en el intervalo $[\pi/4, 3\pi/4]$, la nefroide es la gráfica de otra función distinta $y(x)$, y $x(t)$ es decreciente, luego tenemos también la igualdad

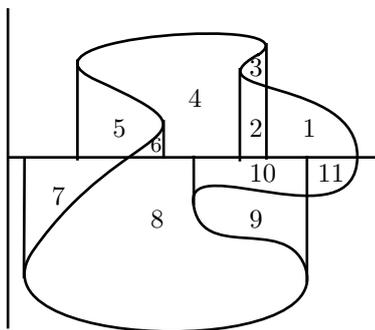
$$-\int_{x(\pi/4)}^{x(3\pi/4)} y(x) dx = -\int_{\pi/4}^{3\pi/4} y(t)x'(t) dt,$$

pero ahora los extremos de la primera integral están invertidos (porque $x(t)$ es decreciente), luego la primera integral da el área cambiada de signo, y con el signo antepuesto se vuelve positiva. Así pues, la integral con el signo antepuesto es el área (positiva) que queda bajo la función $y(x)$, es decir, el área abarcada por la nefroide sobre el eje X incluyendo las dos zonas extra sombreadas en la figura.

Similarmente se razona que la tercera integral es el área sombreada de la izquierda, pero negativa, pues en este caso, el cambio de variable es creciente, como en la primera integral. Por lo tanto, la suma de las tres primeras integrales es exactamente el área abarcada por la nefroide por encima del eje X , ya que la primera y la tercera restan lo que le sobra a la segunda.

Con esto basta para calcular el área de la nefroide, pues basta multiplicar por 2 el resultado hasta aquí, pero en realidad las tres últimas integrales dan el área que falta. En efecto, en la cuarta integral el cambio de variable es decreciente, pero la función $y(x)$ es negativa, luego la integral es positiva (el integrando es negativo, pero con los extremos invertidos) y con el signo antepuesto queda negativa. Por el contrario, la quinta integral es positiva (integrando negativo, extremos en orden y signo antepuesto negativo) y la sexta es negativa de nuevo. Así, la suma de las seis integrales, es decir, la integral completa entre 0 y 2π , es justamente el área de la nefroide.

Pero es interesante observar que este razonamiento es aplicable a cualquier otra curva $r : [a, b] \rightarrow \mathbb{R}^2$ que sea cerrada (es decir, que cumpla $r(a) = r(b)$) y simple (que no se corte a sí misma): al aplicar el cambio de variable $x = x(t)$ cuando $x(t)$ es (de)creciente e $y(t)$ es positiva, nos da el área que queda por debajo con signo positivo (negativo), y cuando $x(t)$ es (de)creciente e $y(t)$ es negativa, nos da el área que queda por encima con signo negativo (positivo), y esto hace que el resultado sea siempre el área encerrada por la curva si la parametrización la recorre en sentido horario (de modo que el área que encierra está a la derecha respecto del sentido de avance) o el opuesto si la recorre en sentido antihorario (y entonces el área encerrada está a la izquierda respecto al sentido de avance). Como hemos recorrido la nefroide en sentido antihorario, por eso hemos necesitado anteponer un signo negativo. Por ejemplo, si la curva es la siguiente:



y la parametrización empieza en el punto con $y = 0$ situado a la derecha en sentido antihorario, las integrales en los intervalos sucesivos donde $x(t)$ es creciente o decreciente nos van dando las áreas siguientes:

$$\begin{aligned} & -(A_1 + A_2) + (A_2 + A_3) - (A_2 + A_3 + A_4 + A_5 + A_6) + (A_5 + A_6) + A_6 - A_7 \\ & \quad - (A_7 + A_8 + A_9 + A_{10}) + (A_9 + A_{10}) - (A_{10} + A_{11}) \\ & = -(A_1 + A_2 + A_4 + A_6 + A_8 + A_{10} + A_{11}), \end{aligned}$$

que es justamente la suma de las áreas de las regiones abarcadas por la curva con signo negativo. Por ejemplo, en el primer tramo y es positiva y $x(t)$ es decreciente, luego la integral da $-(A_1 + A_2)$, etc.

No estamos en condiciones de demostrar que esto no es casualidad,⁵ pero el lector puede comprobar en cada caso en particular que la integral (2.2) (sin el signo negativo para curvas recorridas en sentido horario) proporciona el área de la región encerrada igual que hemos hecho con la nefroide, lo que le garantizará que el uso de la fórmula es correcto.

Notemos que, por simetría, también es válida la fórmula análoga en la que x e y se intercambian, e incluso hay una fórmula equivalente más simétrica que a menudo es más práctica:

Teorema 2.10 Si $r : [a, b] \rightarrow \mathbb{R}^2$ es una curva cerrada que no se corta a sí misma, el área que encierra es

$$A = \left| \int_a^b y(t)x'(t) dt \right| = \left| \int_a^b x(t)y'(t) dt \right| = \frac{1}{2} \left| \int_a^b (x(t)y'(t) - y(t)x'(t)) dt \right|.$$

DEMOSTRACIÓN: Vamos a demostrar que la tercera fórmula es equivalente a la primera. Para ello observamos que

$$\int_a^b y(t)x'(t) dt + \frac{1}{2} \int_a^b (x(t)y'(t) - y(t)x'(t)) dt = \frac{1}{2} \int_a^b (x(t)y'(t) + y(t)x'(t)) dt,$$

⁵Es consecuencia del teorema de Green.

luego basta ver que la última integral es nula. Ahora bien, si integramos por partes:

$$\int_a^b x(t)y'(t) dt = [x(t)y(t)]_a^b - \int_a^b y(t)x'(t) dt = - \int_a^b y(t)x'(t) dt,$$

donde usamos que la curva es cerrada, por lo que $x(a) = x(b)$, $y(a) = y(b)$. ■

En la sección B.3 está calculada el área de la nefroide usando la tercera fórmula del teorema anterior.

Integración de series de potencias En este apartado probaremos el teorema 1.21, cuya demostración habíamos dejado pendiente en el capítulo anterior y que hemos usado para demostrar el teorema 1.22 sobre derivación de series de potencias, y a continuación usaremos este teorema, que ya estará demostrado, para probar un resultado análogo sobre integración de series de potencias.

DEMOSTRACIÓN (de 1.21): El criterio de mayoración de Weierstrass [ITAn 3.7] garantiza que la serie de las derivadas de las funciones dadas converge a una función continua $g :]a, b[\rightarrow \mathbb{R}$, de modo que

$$g(x) = \sum_{n=0}^{\infty} f'_n(x).$$

Definimos

$$f(x) = \sum_{n=0}^{\infty} f_n(x_0) + \int_{x_0}^x g(t) dt,$$

con lo que trivialmente $f' = g$.

Sólo falta probar que, para todo $x \in]a, b[$, se cumple que

$$f(x) = \sum_{n=0}^{\infty} f_n(x)$$

Para ello observamos que

$$f_n(x) = f_n(x_0) + \int_{x_0}^x f'_n(t) dt,$$

por lo que

$$\begin{aligned} \left| f(x) - \sum_{n=0}^k f_n(x) \right| &= \left| \sum_{n=0}^{\infty} f_n(x_0) + \int_{x_0}^x g(t) dt - \sum_{n=0}^k f_n(x_0) - \sum_{n=0}^k \int_{x_0}^x f'_n(t) dt \right| \leq \\ &\left| \sum_{n=k+1}^{\infty} f_n(x_0) \right| + \left| \int_{x_0}^x \sum_{n=k+1}^{\infty} f'_n(t) dt \right| \leq \left| \sum_{n=k+1}^{\infty} f_n(x_0) \right| + \int_{x_0}^x \sum_{n=k+1}^{\infty} |f'_n(t)| dt \leq \\ &\left| \sum_{n=k+1}^{\infty} f_n(x_0) \right| + \int_{x_0}^x \sum_{n=k+1}^{\infty} M_n dt \leq \left| \sum_{n=k+1}^{\infty} f_n(x_0) \right| + \sum_{n=k+1}^{\infty} M_n |x - x_0| \leq \\ &\left| \sum_{n=k+1}^{\infty} f_n(x_0) \right| + (b - a) \sum_{n=k+1}^{\infty} M_n. \end{aligned}$$

Así, dado $\epsilon > 0$, existe un k_0 tal que si $k \geq k_0$ se cumple que

$$\left| \sum_{n=k+1}^{\infty} f_n(x_0) \right| < \frac{\epsilon}{2}, \quad \sum_{n=k+1}^{\infty} M_n < \frac{\epsilon}{2(b-a)},$$

con lo que

$$\left| f(x) - \sum_{n=0}^k f_n(x) \right| < \epsilon,$$

y esto prueba la que la serie dada converge a f . ■

Con esto ya tenemos demostrado el teorema 1.22.

El teorema 2.4 garantiza que cualquier función continua, como $f(t) = e^{t^2}$ tiene primitiva, y hemos visto cómo aproximarla aproximando el área bajo su gráfica. Sin embargo, podemos encontrar una expresión explícita para dicha primitiva teniendo en cuenta que

$$f(t) = \sum_{n=0}^{\infty} \frac{t^{2n}}{n!},$$

por lo que

$$\int f(t) dt = \int \sum_{n=0}^{\infty} \frac{t^{2n}}{n!} dt = \sum_{n=0}^{\infty} \int \frac{t^{2n}}{n!} dt = \sum_{n=0}^{\infty} \frac{t^{2n+1}}{(2n+1)n!} + c.$$

El intercambio de la serie con la integral está justificado por el teorema 1.22 o, más precisamente, por el teorema siguiente, que es una consecuencia inmediata del teorema indicado:

Teorema 2.11 Si $f(z) = \sum_{n=0}^{\infty} a_n z^n$ es una serie de potencias con radio de convergencia $R > 0$, entonces la serie

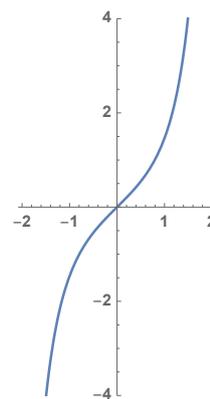
$$F(z) = \sum_{n=1}^{\infty} \frac{a_n}{n+1} z^{n+1} + c$$

tiene el mismo radio de convergencia, la función $F(z)$ es derivable en su disco de convergencia y $F'(z) = f(z)$.

Sucede que la función

$$F(t) = \sum_{n=1}^{\infty} \frac{t^{2n+1}}{(2n+1)n!}$$

no puede expresarse en términos de funciones “usuales” (exponenciales, senos, cosenos, etc.), pero su expresión como serie de potencias es todo lo necesario para que un ordenador pueda tratar con



ella exactamente igual que trata a dichas funciones “usuales”. Si esta función tuviera más interés, nada impediría construir calculadoras de bolsillo que la calcularan igual que calculan senos y cosenos. La figura muestra su gráfica.

Nota Conviene observar que, en principio, tenemos la definición de derivabilidad para funciones de variable real y la definición para funciones de variable compleja. Son formalmente idénticas, pero una se aplica a funciones definidas en un abierto en \mathbb{R} y otra a funciones definidas en un abierto en \mathbb{C} . Ahora bien, si una función está definida en un abierto $A \subset \mathbb{C}$ y $A' = A \cap \mathbb{R} \neq \emptyset$, es fácil ver que A' es un abierto en \mathbb{R} y, si f toma sobre A' valores reales, como le sucede a las funciones e^z , $\operatorname{sen} z$, $\operatorname{cos} z$, o la función e^{t^2} , es inmediato que si f es derivable en A , también lo es en A' como función de variable real, y la derivada es la misma en ambos sentidos, pues la definición de derivada de variable compleja implica la definición de derivada real. ■

2.3 Las leyes de Kepler

En esta sección formularemos matemáticamente las leyes de Kepler que hemos discutido en la introducción, lo que nos permitirá en el capítulo IV demostrarlas a partir de las leyes de Newton. Recordemos la primera:

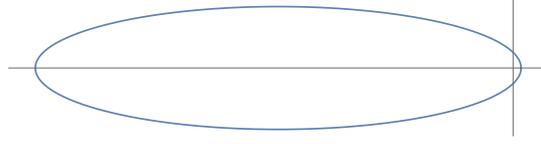
Primera ley de Kepler *Los planetas siguen órbitas elípticas con el Sol en uno de sus focos.*

Más en general, sucede que todo cuerpo que se mueva bajo la acción de la gravedad del Sol, o bien se dirige a él en línea recta hasta estrellarse, o bien se aleja de él en línea recta, o bien sigue una trayectoria cónica, que puede ser una elipse, o una parábola o una hipérbola. Ahora bien, en los dos últimos casos, la trayectoria es abierta y sólo se acerca al Sol una vez, por lo que un planeta que orbite periódicamente alrededor del Sol debe seguir necesariamente una trayectoria elíptica, tal y como Kepler afirmó.

En la discusión previa al teorema 1.37 vimos que la ecuación de una cónica en coordenadas polares respecto de un sistema de referencia con origen en uno de sus focos es de la forma

$$\rho = \frac{l}{1 + \epsilon \cos(\theta + \phi)},$$

donde $\epsilon \geq 0$ es la excentricidad de la cónica. Por ejemplo, el cometa Halley describe una elipse de excentricidad $\epsilon = 0.968$, por lo que su órbita tiene el aspecto que muestra la figura siguiente, donde hemos tomado $l = 1$, pero esto sólo afecta a la escala. Aunque pusiéramos su valor real, al representar la órbita a una escala adecuada para que cupiera en la página, obtendríamos la misma figura.



Observemos que

$$\rho(0) = \frac{l}{1 + \epsilon}, \quad \rho(\pi) = \frac{l}{1 - \epsilon}$$

corresponden, respectivamente, al *perihelio* del cometa, es decir, a la distancia mínima al Sol y, en el caso de una órbita elíptica, al *afelio* o punto de la órbita más alejado del Sol. Por otra parte, $\rho(\pi/2) = l$, por lo que $2l$ se interpreta geoméricamente como el *latus rectum* de la cónica, es decir, la longitud de la cuerda perpendicular al eje mayor que pasa por un foco. Para una órbita elíptica, el semieje mayor es

$$a = \frac{\rho_0 + \rho_\pi}{2} = \frac{l}{1 - \epsilon^2}$$

y el semieje menor, según se ve en la sección 7.1 de [IGE], es $b = \sqrt{a^2 - c^2}$, con $c = \epsilon a$, luego

$$b = a\sqrt{1 - \epsilon^2} = \frac{l}{\sqrt{1 - \epsilon^2}}.$$

Con esto sabemos dónde podemos encontrar un planeta o un cometa, etc. en un momento dado y dónde no, pero la primera ley de Kepler no nos informa sobre la velocidad a la que se mueve en su órbita. Esto lo determina la segunda ley:

Segunda ley de Kepler *El radio que une un planeta con el Sol barre áreas iguales en tiempos iguales.*

Para traducir esta ley en una fórmula explícita usamos el teorema 2.6, según el cual, el área barrida por el radio que une el Sol con el cuerpo entre dos ángulos θ_0 y θ_1 viene dada por

$$A = \frac{1}{2} \int_{\theta_0}^{\theta_1} \rho^2 d\theta = \frac{l^2}{2} \int_{\theta_0}^{\theta_1} \frac{1}{(1 + \epsilon \cos \theta)^2} d\theta.$$

Esta integral la hemos calculado en (C.16). Por simplificar, calculamos el área entre $\theta_0 = 0$ y $\theta_1 = \theta$, que resulta ser

$$A(\theta) = \frac{l^2}{(1 - \epsilon^2)\sqrt{1 - \epsilon^2}} \arctan \left(\sqrt{\frac{1 - \epsilon}{1 + \epsilon}} \frac{\text{sen } \theta}{1 + \cos \theta} \right) - \frac{l^2}{2(1 - \epsilon^2)} \frac{\epsilon \text{ sen } \theta}{1 + \epsilon \cos \theta},$$

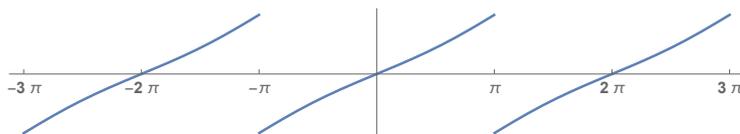
aunque podemos simplificar un poco la expresión de las constantes:

$$A(\theta) = ab \arctan \left(\sqrt{\frac{1 - \epsilon}{1 + \epsilon}} \frac{\text{sen } \theta}{1 + \cos \theta} \right) - \frac{b^2}{2} \frac{\epsilon \text{ sen } \theta}{1 + \epsilon \cos \theta}.$$

Ahora bien, aquí tenemos que hacer una reflexión. La función

$$A(\theta) = \frac{l^2}{2} \int_0^\theta \frac{1}{(1 + \epsilon \cos t)^2} dt$$

es monótona creciente, mientras que la expresión explícita que hemos encontrado es periódica de periodo 2π . Por ejemplo, su gráfica para la excentricidad de Marte es



Ello se debe a que la hemos calculado con el cambio de variable $x = \tan(\theta/2)$, que es biyectivo como aplicación $x :]-\pi, \pi[\rightarrow \mathbb{R}$, por lo que la primitiva que hemos calculado sólo vale en el intervalo $]-\pi, \pi[$, pero esto basta para hacer cálculos.

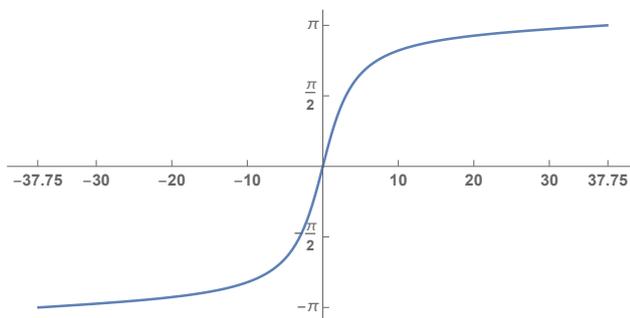
Por ejemplo, el cometa Halley da una vuelta alrededor del Sol cada 75.3 años, lo que significa que su *velocidad areolar*, es decir, el área que barre por unidad de tiempo es de

$$v_a = \frac{\pi ab}{75.3} \text{ unidades de área/año,}$$

luego se encontrará en la posición θ de su órbita en el instante (en años)

$$t(\theta) = \frac{75.3}{\pi ab} A(\theta) = \frac{75.3}{\pi} \arctan \left(\sqrt{\frac{1-\epsilon}{1+\epsilon}} \frac{\sin \theta}{1 + \cos \theta} \right) - \frac{75.3b}{2a\pi} \frac{\epsilon \sin \theta}{1 + \cos \theta}.$$

La figura siguiente muestra la gráfica de la función $\theta(t)$ (es simplemente la gráfica de $t(\theta)$ representada con θ en el eje vertical). En ella vemos cómo la segunda ley de Kepler determina dónde se encuentra el cometa Halley en cada instante. No se puede obtener una expresión explícita para $\theta(t)$, pero se puede aproximar numéricamente su valor para cualquier t .



Por ejemplo, en la figura siguiente hemos dividido la órbita del cometa en 76 intervalos recorridos en tiempos iguales (lo que corresponde a un tiempo de 0.99 años por intervalo):



Por completitud citamos aquí la tercera ley de Kepler, aunque su formulación matemática es obvia y no tenemos nada que añadir:

Tercera ley de Kepler *Los cuadrados de los períodos de revolución de los planetas son proporcionales a los cubos de los semiejes mayores de sus órbitas.* ■

2.4 El teorema de aproximación de Weierstrass

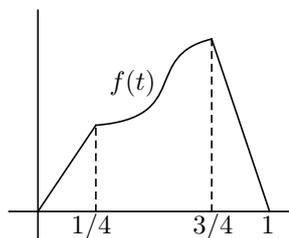
Vamos a ver cómo el cálculo integral nos permite demostrar un enunciado que, en principio, nada tiene que ver con integrales:

Teorema 2.12 (Teorema de aproximación de Weierstrass) *Dada una función continua $f : [a, b] \rightarrow \mathbb{R}$ y $\epsilon > 0$, existe un polinomio $p(x)$ tal que, para todo $x \in [a, b]$, se cumple que $|f(x) - p(x)| < \epsilon$.*

DEMOSTRACIÓN: Observamos en primer lugar que no perdemos generalidad si suponemos que $a = 1/4$ y $b = 3/4$, pues para otros valores $a < b$ podemos considerar una aplicación $h : [1/4, 3/4] \rightarrow [a, b]$ de la forma⁶ $h(x) = u + xv$ que haga corresponder ambos intervalos, con inversa $h^{-1}(t) = (t - u)/v$. Así, la función $g(x) = f(h(x))$ es continua en $[1/4, 3/4]$ y, si probamos el teorema para este intervalo, tendremos la existencia de un polinomio $p_0(x)$ tal que $|f(h(x)) - p_0(x)| < \epsilon$, para todo $x \in [1/4, 3/4]$, luego también se cumple que $|f(t) - p_0(h^{-1}(t))| < \epsilon$ para todo $t \in [a, b]$, y la función $p(t) = p_0(h^{-1}(t))$ es un polinomio que cumple lo requerido.

Así, podemos suponer que $f : [1/4, 3/4] \rightarrow \mathbb{R}$ es una función continua, que a su vez podemos extender a una función continua en \mathbb{R} mediante

$$F(t) = \begin{cases} 0 & \text{si } t \leq 0 \text{ o } t \geq 1, \\ 4f(a)t & \text{si } 0 \leq t \leq 1/4, \\ f(t) & \text{si } 1/4 \leq t \leq 3/4, \\ 4f(b)(1-t) & \text{si } 3/4 \leq t \leq 1. \end{cases}$$



⁶Explícitamente, $h(x) = 2(b-a)x + (3a-b)/2$, pero los valores concretos de u y v son irrelevantes.

Si aproximamos con un polinomio la función F en el intervalo $[0, 1]$, en particular dicho polinomio aproximará a f con la misma precisión en el intervalo $[1/4, 3/4]$, luego podemos partir de una función continua $f : [0, 1] \rightarrow \mathbb{R}$ con la condición adicional de que $f(0) = f(1) = 0$, lo que nos permite extenderla a una función continua $f : \mathbb{R} \rightarrow \mathbb{R}$ con el valor 0 fuera del intervalo $[0, 1]$.

Vamos a probar que las funciones

$$p_n(x) = \frac{1}{c_n} \int_{-1}^1 f(t)(1 - (t - x)^2)^n dt,$$

donde

$$c_n = \int_{-1}^1 (1 - t^2)^n dt,$$

son polinomios que aproximan a $f(x)$ con precisión mayor cuanto mayor es n . El valor de c_n lo hemos calculado en (2.1), pero no lo vamos a necesitar. Nos bastará con la estimación $c_n \geq 2/(n+1)$, que podemos obtener fácilmente. Para ello observamos que la integral que define a c_n se puede descomponer en suma de la integral en $[-1, 0]$ más la integral en $[0, 1]$, y que el cambio de variable $t = -t'$ transforma la primera en la segunda, por lo que

$$\begin{aligned} c_n &= \int_{-1}^1 (1 - t^2)^n dt = 2 \int_0^1 (1 - t^2)^n dt = 2 \int_0^1 (1 - t)^n (1 + t)^n dt \geq \\ &2 \int_0^1 (1 - t)^n dt = -2 \left[\frac{(1 - t)^{n+1}}{n+1} \right]_0^1 = \frac{2}{n+1}. \end{aligned}$$

Que $p_n(x)$ es un polinomio es fácil de ver, pues el factor $(1 - (t - x)^2)^n$ es ciertamente un polinomio en dos variables, que podemos expresar como

$$(1 - (t - x)^2)^n = \sum_{i=0}^{2n} q_i(t)x^i,$$

para ciertos polinomios $q_i(t)$. Por consiguiente:

$$p_n(x) = \frac{1}{c_n} \int_{-1}^1 \sum_{i=0}^{2n} q_i(t)f(t)x^i dt = \sum_{i=0}^{2n} \left(\frac{1}{c_n} \int_{-1}^1 q_i(t)f(t) dt \right) x^i.$$

Así, pues, en efecto, $p_n(x)$ es un polinomio. Ahora fijamos $0 \leq x \leq 1$ y observamos que

$$p_n(x) = \frac{1}{c_n} \int_{-1+x}^{1+x} f(t)(1 - (t - x)^2)^n dt,$$

pues f se anula en los intervalos $[-1 + x, 0]$ y $[1, 1 + x]$. Ahora, el cambio de variable $t' = t - x$ transforma esta expresión en

$$p_n(x) = \frac{1}{c_n} \int_{-1}^1 f(x+t)(1 - t^2)^n dt = \int_{-1}^1 f(x+t)K_n(t) dt,$$

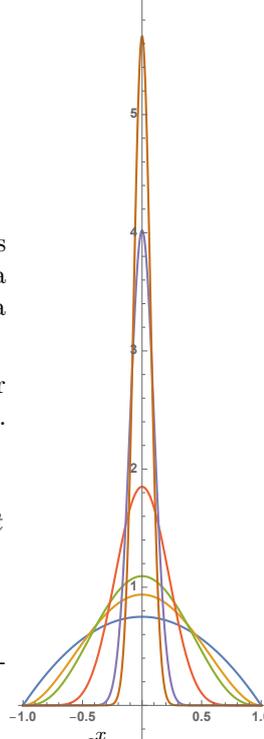
donde hemos introducido lo que se conoce como el *núcleo de Landau*:

$$K_n(t) = \frac{1}{c_n}(1-t^2)^n.$$

Por la definición de c_n tenemos que

$$\int_{-1}^1 K_n(t) dt = 1.$$

La figura muestra las funciones $K_n(t)$ para los valores $n = 1, 2, 3, 10, 50, 100$. Vemos que la gráfica se hace cada vez más estrecha y, a la vez, más alta para que el área que limita sea siempre igual a 1.



Vamos a probar que, en efecto, el área limitada por la gráfica se concentra mayoritariamente alrededor de 0. Para ello fijamos $0 < \delta < 1$ y observamos que

$$\begin{aligned} \int_{\delta}^1 K_n(t) dt &= \frac{1}{c_n} \int_{\delta}^1 (1-t^2)^n dt \leq \frac{n+1}{2} \int_{\delta}^1 (1-\delta)^n dt \\ &= \frac{1}{2}(1-\delta)(1-\delta^2)^n(n+1) \end{aligned}$$

y la última expresión tiende a 0 cuando n tiende a infinito, por ejemplo porque, si $0 < c < 1$,

$$\lim_{x \rightarrow +\infty} c^x(x+1) = \lim_{x \rightarrow +\infty} \frac{x+1}{c^{-x}} = \lim_{x \rightarrow +\infty} -\frac{1}{c^{-x} \log c} = \lim_{x \rightarrow +\infty} -\frac{c^x}{\log c} = 0.$$

Por el teorema de Weierstrass [ITAn 3.31] existe un M tal que $|f(t)| \leq M$ para todo $t \in [0, 1]$, luego, de hecho, para todo $t \in \mathbb{R}$. El teorema de Heine-Cantor [ITAn A.14] aplicado a f sobre el intervalo $[-1, 2]$ nos da la existencia de un $0 < \delta < 1$ tal que si $x \in [0, 1]$ y $|t| < \delta$, entonces

$$|f(x) - f(x+t)| < \frac{\epsilon}{2}$$

Por consiguiente,

$$\begin{aligned} |f(x) - p_n(x)| &= \left| f(x) \int_{-1}^1 K_n(t) dt - \int_{-1}^1 f(x+t)K_n(t) dt \right| \leq \\ &\int_{-1}^1 |f(x) - f(x+t)|K_n(t) dt = \int_{-1}^{-\delta} |f(x) - f(x+t)|K_n(t) dt + \\ &\int_{-\delta}^{\delta} |f(x) - f(x+t)|K_n(t) dt + \int_{\delta}^1 |f(x) - f(x+t)|K_n(t) dt < \\ &\frac{\epsilon}{2} \int_{-1}^1 K_n(t) dt + 2M \int_{-1}^{-\delta} K_n(t) dt + 2M \int_{\delta}^1 K_n(t) dt = \frac{\epsilon}{2} + 4M \int_{\delta}^1 K_n(t) dt, \end{aligned}$$

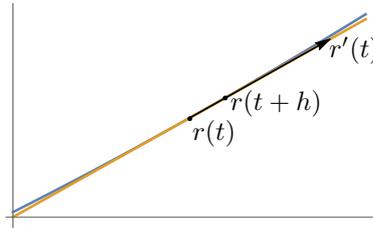
donde en el último paso hemos usado que $K_n(t) = K_n(-t)$, por lo que el cambio de variable $t = -t'$ transforma la segunda integral en la tercera. Y hemos probado que la última integral tiende a 0 cuando n tiende a infinito, luego existe un n_0 tal que si $n \geq n_0$, entonces la integral es menor que $\epsilon/8M$, lo que nos da que $|f(x) - p_n(x)| < \epsilon$, como había que probar. ■

2.5 La longitud de un arco

Un *arco* es una función $r : [a, b] \rightarrow \mathbb{R}^3$, que supondremos derivable con derivada continua. Usamos la palabra “arco” en lugar de “curva” para enfatizar que tiene un *extremo inicial* $r(a)$ y un *extremo final* $r(b)$.

Vamos a asociarle otra función $s : [a, b] \rightarrow \mathbb{R}$ de modo que, si pensamos en r como la trayectoria de un punto que se mueve desde $r(a)$ hasta $r(b)$, entonces $s(t)$ es la longitud recorrida por el punto hasta el instante t , pero medida sobre la trayectoria (por ejemplo, si r es un arco de circunferencia, entonces $s(t)$ será la longitud del arco desde $r(a)$ hasta $r(t)$, no la longitud de la cuerda que une ambos puntos en línea recta).

Si ampliamos suficientemente la gráfica alrededor de un punto $r(t)$ veremos algo así:



El punto $r(t+h)$ es indistinguible del punto $r(t) + r'(t)h$, luego el espacio recorrido sobre el arco entre los instantes t y $t+h$, es decir, $s(t+h) - s(t)$ resulta indistinguible del espacio recorrido sobre la recta tangente, que es $\|r'(t)\| |h|$. Teniendo en cuenta que la función $s(t)$ (que queremos definir) tiene que ser creciente, podemos escribir

$$s(t+h) - s(t) \approx \|r'(t)\| h,$$

pues ambos miembros serán positivos o negativos según si h es positivo o negativo. Como el parecido entre el arco y su recta tangente es mayor cuanto más ampliamos la gráfica (o cuanto menor tomamos h), podemos afirmar que la aproximación

$$\frac{s(t+h) - s(t)}{h} \approx \|r'(t)\|$$

tiene que ser mayor cuanto menor sea h , lo que se traduce en que

$$\frac{ds}{dt} = \|r'(t)\|.$$

Así pues, este argumento apunta a que el módulo de la velocidad de un móvil es la derivada del espacio $s(t)$ medido sobre la trayectoria. Vamos a adoptar esto como definición y a continuación justificaremos que tal definición es razonable.

Definición 2.13 Sea $r : [a, b] \rightarrow \mathbb{R}^3$ una función derivable en $]a, b[$ tal que $r'(t)$ se extiende a una función continua en $[a, b]$. En tal caso diremos que r es *rectificable* en $[a, b]$ y definimos $s : [a, b] \rightarrow \mathbb{R}$ mediante

$$s(t) = \int_a^t \|r'(t)\| dt.$$

En particular, definimos

$$L(r) = \int_a^b \|r'(t)\| dt.$$

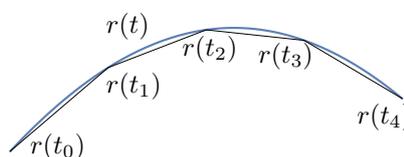
Notemos que la condición se cumple en particular si r puede extenderse a un arco definido en un intervalo abierto que contenga a $[a, b]$ en el que tenga derivada continua.

El teorema siguiente nos da una interpretación precisa del valor $L(r)$ que acabamos de definir:

Teorema 2.14 Si $r : [a, b] \rightarrow \mathbb{R}^3$ es un arco rectificable, para todo $\epsilon > 0$ existe un $\delta > 0$ tal que si $a = t_0 < t_1 < \dots < t_m = b$ es una partición de $[a, b]$ en intervalos de longitud $\Delta t_i = t_i - t_{i-1} < \delta$, entonces

$$\left| \sum_{i=1}^m \|r(t_i) - r(t_{i-1})\| - L(r) \right| < \epsilon.$$

Esto significa que $L(r)$ se puede aproximar por la suma de las longitudes de los segmentos de una poligonal que una varios puntos de la curva, y la aproximación es mejor cuanto más próximos estén los puntos.



DEMOSTRACIÓN: Por el teorema 2.5, existe un $\delta > 0$ tal que si

$$a = t_0 < t_1 < \dots < t_m = b$$

es una partición de $[a, b]$ en intervalos de longitud $t_i - t_{i-1} < \delta$, entonces

$$\left| \sum_{i=1}^m \|r'(t_{i-1})\| \Delta t_i - \int_a^b \|r'(t)\| dt \right| < \frac{\epsilon}{2}.$$

Por el teorema de Heine-Cantor [ITAn A.14] aplicado a las derivadas r'_j de las funciones coordenadas de la función r (que por definición de arco rectificable se extienden a funciones continuas en $[a, b]$), podemos suponer que si $|t_1 - t_2| < \delta$, entonces

$$|r'_j(t_1) - r'_j(t_2)| < \frac{\epsilon}{2\sqrt{3}(b-a)}.$$

Consideremos una partición que cumpla $\Delta t_i < \delta$ para $i = 1, \dots, m$. Ahora aplicamos el teorema del valor medio a cada una de las funciones coordenadas r_j de la función r , según el cual existen puntos $t_{i-1} < t_{ij} < t_i$ tales que

$$r_j(t_i) - r_j(t_{i-1}) = r'_j(t_{ij})\Delta t_i.$$

Como $|t_{ij} - t_{i-1}| \leq \Delta t_i < \delta$, tenemos que

$$|r'_j(t_{ij}) - r'_j(t_{i-1})| < \frac{\epsilon}{2\sqrt{3}(b-a)}.$$

Ahora usamos la forma siguiente de la desigualdad triangular:

$$|\|x\| - \|y\|| \leq \|x - y\|,$$

y la aplicamos a los vectores $r(t_i) - r(t_{i-1})$ y a $r'(t_{i-1})\Delta t_i$. La conclusión es que

$$\begin{aligned} & |\|r(t_i) - r(t_{i-1})\| - \|r'(t_{i-1})\Delta t_i\| \leq \|r(t_i) - r(t_{i-1}) - r'(t_{i-1})\Delta t_i\| \\ &= \sqrt{\sum_{j=1}^3 (r'_j(t_{ij})\Delta t_i - r'_j(t_{i-1})\Delta t_i)^2} = \sqrt{\sum_{j=1}^3 |r'_j(t_{ij}) - r'_j(t_{i-1})|^2 \Delta t_i} \\ &\leq \sqrt{\sum_{j=1}^3 \frac{\epsilon^2}{12(b-a)^2} \Delta t_i} = \sqrt{\frac{\epsilon^2}{4(b-a)^2} \Delta t_i} = \frac{\epsilon}{2(b-a)} \Delta t_i. \end{aligned}$$

Por consiguiente:

$$\begin{aligned} & \left| \sum_{i=1}^m \|r(t_i) - r(t_{i-1})\| - \sum_{i=1}^m \|r'(t_{i-1})\Delta t_i\| \right| \leq \\ & \sum_{i=1}^m |\|r(t_i) - r(t_{i-1})\| - \|r'(t_{i-1})\Delta t_i\|| \leq \frac{\epsilon}{2(b-a)} \sum_{i=1}^m \Delta t_i = \frac{\epsilon}{2}, \end{aligned}$$

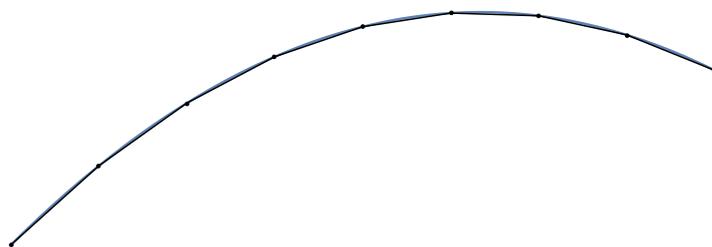
y en conclusión:

$$\begin{aligned} & \left| \sum_{i=1}^m \|r(t_i) - r(t_{i-1})\| - L(r) \right| \leq \\ & \left| \sum_{i=1}^m \|r(t_i) - r(t_{i-1})\| - \sum_{i=1}^m \|r'(t_{i-1})\Delta t_i\| \right| + \left| \sum_{i=1}^m \|r'(t_{i-1})\Delta t_i\| - L(r) \right| < \epsilon. \end{aligned}$$

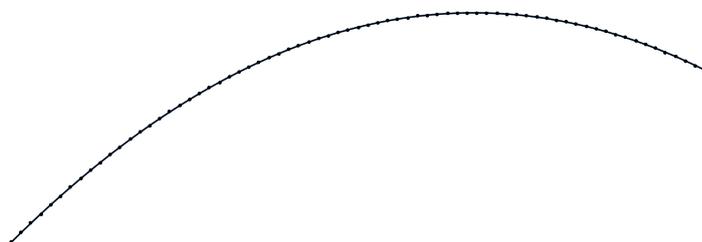
Analicemos el significado de este teorema: imaginemos que $r(t)$ es la posición en el instante t de una partícula puntual y, fijada una partición t_i de $[a, b]$, consideremos otra partícula puntual que, entre el instante t_{i-1} y el instante t_i se mueve de $r(t_{i-1})$ a $r(t_i)$, no a través del arco r , sino en línea recta, a través del segmento con dichos extremos. Entonces, la suma

$$\sum_{i=1}^m \|r(t_i) - r(t_{i-1})\|$$

es exactamente el espacio que recorre esta segunda partícula. Pero si tomamos una partición con muchos puntos muy próximos entre sí, las trayectorias de ambas partículas serán prácticamente indistinguibles:



De hecho, en la figura hemos tomado un número de puntos suficientemente pequeño como para que se aprecie que hay cierta diferencia, porque si lo tomamos mayor ya no se nota la diferencia:



El teorema anterior prueba que, cuando las distancias Δt_i se hacen cada vez menores, la longitud de la poligonal converge a $L(r)$, por lo que es natural considerar que $L(r)$ es el espacio recorrido por una partícula puntual que recorra el arco.

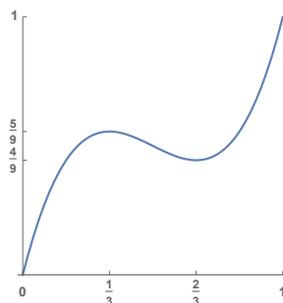
En este punto es importante destacar un detalle: si el arco r es inyectivo, es decir, si valores distintos del parámetro t se corresponden con puntos distintos $r(t)$, entonces podemos decir que $L(r)$ es la *longitud* del arco r , pero si la parametrización de r hace que la partícula “se pare y vuelva sobre sus pasos” o que vuelva al punto de partida y entre en un ciclo, hay que tener en cuenta que $L(r)$ mide el espacio recorrido, de modo que la longitud de cada tramo se contará en $L(r)$ tantas veces como se recorra.

Ejemplo Consideremos la función

$$f(t) = 6t^3 - 9t^2 + 4t.$$

Es fácil ver que se comporta como muestra la gráfica siguiente. Vemos que sube desde 0 hasta $5/9$, luego desciende hasta $4/9$ y vuelve a subir hasta 1. Ahora consideremos el arco $r : [0, 1] \rightarrow \mathbb{R}^2$ dado por

$$r(t) = (\cos(\pi f(t)), \operatorname{sen}(\pi f(t))).$$



Su gráfica es simplemente una semicircunferencia de radio 1, pero no es cierto que $L(r) = \pi$, sino que $L(r) = 11\pi/9$, porque una partícula puntual que en el instante t se encuentre en el punto $r(t)$ avanza un arco de longitud $5\pi/9$, luego retrocede un arco de longitud $\pi/9$, y luego vuelve a avanzar un arco de longitud $5\pi/9$, luego la distancia total que recorre es $11\pi/9$. El cálculo explícito no es complicado:

$$r'(t) = (-\pi f'(t) \operatorname{sen}(\pi f(t)), -\pi f'(t) \operatorname{cos}(\pi f(t))),$$

luego $\|r'(t)\| = \pi\sqrt{f'(t)^2} = \pi|f'(t)|$. Por consiguiente, teniendo en cuenta que $f'(t)$ es negativa en $[1/3, 2/3]$,

$$\begin{aligned} L(r) &= \pi \int_0^1 |f'(t)| dt = \pi \int_0^{1/3} f'(t) dt - \pi \int_{1/3}^{2/3} f'(t) dt + \pi \int_{2/3}^1 f'(t) dt \\ &= \pi(f(1/3) - f(0) - f(2/3) + f(1/3) + f(1) - f(2/3)) \\ &= \pi(f(1) - f(0) + 2f(1/3) - 2f(2/3)) = \pi(1 - 0 + 10/9 - 8/9) = 11\pi/9. \end{aligned}$$

■

En general, nos referiremos a $L(r)$ como la *longitud* del arco r , pero hay que tener presente que sólo es la longitud propiamente dicha si la parametrización considerada no recorre dos veces un mismo tramo.

Conviene observar que la longitud es invariante por cambios de parámetro en el sentido siguiente:

Definición 2.15 Dado un arco parametrizable $r : [a, b] \rightarrow \mathbb{R}^3$, un *cambio de parámetro* es una aplicación $t : [u, v] \rightarrow [a, b]$ biyectiva, continua y derivable en $]u, v[$ con derivada continua no nula. En virtud del teorema 1.30, la derivada t' tiene signo constante (o bien es siempre positiva, o bien es siempre negativa), lo cual se traduce en que t es estrictamente creciente o bien estrictamente decreciente en $[u, v]$.

Entonces $r(s) = r(t(s))$ define un arco parametrizado rectificable sobre $[u, v]$, pues, por la regla de la cadena aplicada a cada una de sus componentes $r_i(t(s))$, es derivable con derivada $r'(t(x))t'(s)$. Se dice que $r(s)$ es la *reparametrización* de $r(t)$ por el cambio de parámetro $t(s)$.

Lo que afirmábamos antes de la definición es que $L(r)$ es el mismo valor tanto si se calcula a partir de $r(t)$ o de la reparametrización $r(s)$. En efecto, el cálculo con la reparametrización es

$$\int_u^v \|r'(t(s))t'(s)\| ds = \pm \int_u^v \|r'(t(s))\| |t'(s)| ds = \int_a^b \|r'(t)\| dt,$$

donde hemos usado el teorema de cambio de variable (y hemos tenido en cuenta que si $t' < 0$ es que el cambio de parámetro es decreciente, luego al hacer el cambio de variable quedará la integral desde b hasta a con signo negativo, el cual desaparece al invertir los extremos de la integral).

Por ejemplo, en [ITAn 5.7] probamos que un arco de amplitud θ en una circunferencia de radio 1 tiene longitud θ , y obtuvimos este valor aproximando el arco por la poligonal que resulta de dividir el ángulo en n partes iguales y haciendo tender n a ∞ . El teorema anterior implica que la longitud así calculada es la misma que hemos definido en 2.13, pero ahora es muy fácil calcularla directamente:

Ejemplo Podemos parametrizar un arco de circunferencia como

$$f(t) = (r \cos t, r \sin t),$$

con lo que $f'(t) = (-r \sin t, r \cos t)$ y $\|f'(t)\| = r$. Por consiguiente, la longitud de un arco de radio r y amplitud θ es

$$\int_0^\theta r dt = r\theta.$$

En particular, la longitud de una circunferencia de radio r es $2\pi r$. ■

En general, la longitud de un arco determinado por la gráfica de una función $f(x)$ entre dos valores de x , digamos a y b , puede calcularse usando la parametrización $r(t) = (x, f(x))$, de modo que

$$L = \int_a^b \sqrt{1 + f'(x)^2} dx.$$

Ejemplo: La longitud de una parábola Vamos a calcular la longitud del arco de la parábola $y = ax^2$ comprendido entre $x = 0$ y $x = c > 0$. Según la fórmula anterior, es

$$L = \int_0^c \sqrt{1 + 4ax^2} dx = \frac{1}{2a} \int_0^{2ac} \sqrt{1 + t^2} dt.$$

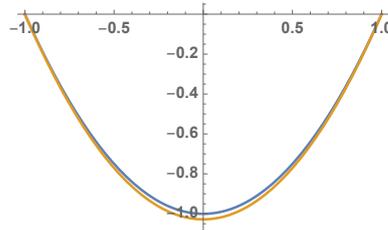
Esta integral está calculada en (C.8):

$$L = \frac{c}{2} \sqrt{1 + 4a^2c^2} + \frac{1}{4a} \operatorname{arsenh} 2ac.$$

En particular, haciendo $a = c = 1$, tenemos que el arco de parábola $y = x^2$ comprendido entre $x = \pm 1$ tiene longitud $2L \approx 2.96$. Si queremos que la longitud sea exactamente 3 tenemos que despejar a en la ecuación

$$\sqrt{1 + 4a^2} + \frac{1}{2a} \operatorname{arsenh} 2a = 3.$$

No es posible hacerlo algebraicamente, pero es fácil aproximar numéricamente la solución: $a \approx 1.027675$. La figura siguiente muestra las parábolas $y = x^2 - 1$ e $y = ax^2 - a$, que pasan ambas por los puntos $(\pm 1, 0)$:



El lector puede ver el cálculo de la longitud de la cicloide, la nefroide y la tractriz en las secciones correspondientes del apéndice B, mientras que para calcular la longitud de la cardioide conviene obtener primero una fórmula para la longitud de una curva expresada en coordenadas polares. La longitud de la lemniscata todavía no sabemos calcularla.

La longitud en coordenadas polares Consideremos una curva de coordenadas polares $(\rho(t), \theta(t))$, es decir, de la curva

$$(x(t), y(t)) = (\rho(t) \cos \theta(t), \rho(t) \operatorname{sen} \theta(t)).$$

Su derivada es

$$T(t) = (\rho' \cos \theta - \rho \operatorname{sen} \theta \theta', \rho' \operatorname{sen} \theta + \rho \cos \theta \theta')$$

y la norma, tras algunas simplificaciones obvias, es

$$\|T(t)\| = \sqrt{\rho'^2 + \rho^2 \theta'^2},$$

luego la longitud de un arco expresado en coordenadas polares resulta ser

$$L = \int_a^b \sqrt{\rho'^2 + \rho^2 \theta'^2} dt. \quad (2.3)$$

Una aplicación de esta fórmula es el cálculo de la longitud de la cardioide en la sección B.2. ■

Ejemplo: La espiral logarítmica Consideremos el caso de la espiral logarítmica (véase la introducción) cuya ecuación en coordenadas polares es

$$\rho = ae^{k\theta},$$

luego $\rho' = ake^{k\theta}$ y la fórmula precedente nos da que

$$L = \int_{\theta_0}^{\theta} \sqrt{a^2e^{2kt} + a^2k^2e^{2kt}} dt = a\sqrt{1+k^2} \int_{\theta_0}^{\theta} e^{kt} dt = a\frac{\sqrt{1+k^2}}{k} [e^{kt}]_{\theta_0}^{\theta},$$

luego concluimos que la longitud de un arco de espiral logarítmica es

$$a\frac{\sqrt{1+k^2}}{k}(e^{k\theta} - e^{k\theta_0}) = \frac{\sqrt{1+k^2}}{k}(\rho - \rho_0).$$

Observemos que la espiral está definida para todo θ real, de modo que no tiene principio ni fin, pero, también podemos decir que la espiral empieza en $(0, 0)$, en el sentido de que

$$\lim_{\theta \rightarrow -\infty} (x(\theta), y(\theta)) = (0, 0).$$

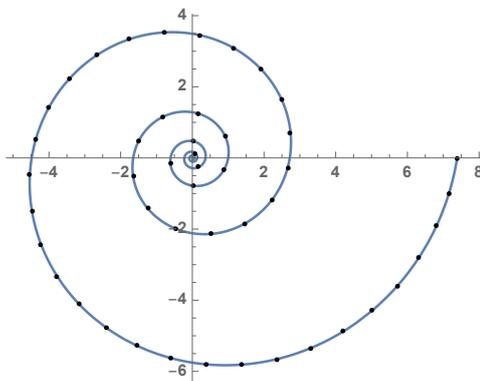
Toricelli se dio cuenta de que la longitud $s(\theta)$ de la espiral “desde su origen” hasta un ángulo θ es finita, en el sentido de que podemos calcular

$$s(\theta) = \lim_{\theta_0 \rightarrow -\infty} a\frac{\sqrt{1+k^2}}{k}(e^{k\theta} - e^{k\theta_0}) = a\frac{\sqrt{1+k^2}}{k}e^{k\theta} = \frac{\sqrt{1+k^2}}{k}\rho.$$

Si despejamos

$$\theta(s) = \frac{1}{k} \log \frac{ks}{a\sqrt{1+k^2}},$$

podemos calcular los parámetros $\theta(1), \theta(2), \theta(3), \dots$ que “midan” la espiral. La figura está calculada con $a = 1$, $k = 1/2\pi$, y vemos que el fragmento representado mide poco más de 47 unidades.



Concretamente, toca el eje de abscisas por última vez en $\rho = e^2$, luego la longitud hasta ese punto es

$$\frac{\sqrt{1+k^2}}{k}\rho = 47.0111\dots$$

■

Parametrización por el arco Un arco parametrizado $r : [a, b] \rightarrow \mathbb{R}^3$ es *regular* si su derivada no se anula en ningún punto. Entonces, la función

$$s(t) = \int_a^t \|r'(u)\| \, du,$$

es una primitiva de $\|r'(t)\|$, es decir, que:

$$\frac{ds}{dt} = \|r'(t)\| > 0,$$

luego $s(t)$ es continua y estrictamente creciente, por lo que su imagen es el intervalo $[0, s(b)]$ y $s : [a, b] \rightarrow [0, s(b)]$ está en las condiciones del teorema de la función inversa, según el cual la función inversa $t : [0, s(b)] \rightarrow [a, b]$ es derivable y su derivada no se anula, pues es

$$t'(s) = \frac{1}{s'(t(s))} = \frac{1}{\|r'(t(s))\|}.$$

Por consiguiente, podemos usar $t(s)$ como cambio de parámetro para obtener una reparametrización $r(s)$ que cumple $r'(s) = r'(t(s))t'(s)$, luego

$$\|r'(s)\| = \|r'(t(s))\| \frac{1}{\|r'(t(s))\|} = 1.$$

En general, si un arco parametrizado regular $r(s)$ cumple $\|r'(s)\| = 1$ para todo valor de s , se dice que está *parametrizado por (la longitud de) el arco*, y esto se traduce en que la longitud del arco comprendido entre dos valores del parámetro s_1 y s_2 es

$$\int_{s_1}^{s_2} \|r'(s)\| \, ds = \int_{s_1}^{s_2} ds = s_2 - s_1,$$

es decir, que aumentar Δs unidades el parámetro equivale a avanzar Δs unidades sobre la curva.

Hemos demostrado que todo arco parametrizado regular puede parametrizarse por el arco mediante un cambio de parámetro adecuado (el dado por el inverso de la función $s(t)$). En general, las parametrizaciones por el arco no se pueden calcular explícitamente, pero su existencia tiene un gran valor teórico.

Ejemplo: La espiral de Arquímedes Consideremos por ejemplo la *espiral de Arquímedes*, dada por

$$r(t) = (t \cos t, t \sin t),$$

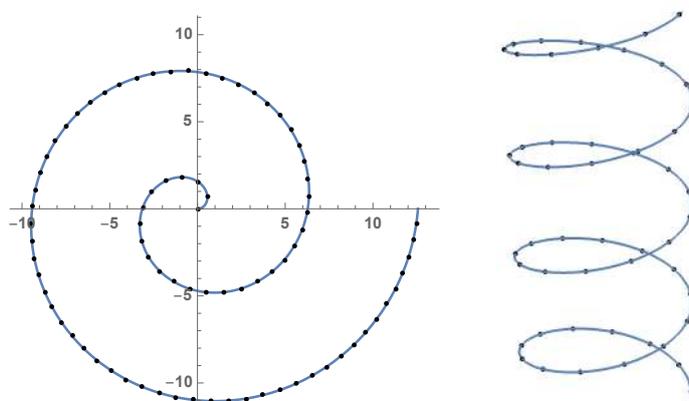
es decir, la trayectoria de una partícula puntual que se aleja de un centro al mismo tiempo que gira a su alrededor, de modo que la velocidad de giro coincide con la velocidad de alejamiento. Entonces

$$r'(t) = (\cos t - t \sin t, \sin t + t \cos t), \quad \|r'(t)\| = \sqrt{t^2 + 1}.$$

Según (C.8),

$$s(t) = \int_0^t \|r'(u)\| du = \int_0^t \sqrt{u^2 + 1} du = \frac{t}{2} \sqrt{1+t^2} + \frac{1}{2} \operatorname{arsenh} t.$$

No es posible calcular explícitamente la función inversa $t(s)$, pero, para cada valor particular de s , sí que es posible resolver numéricamente (es decir, aproximar arbitrariamente la solución de) la ecuación $s(t) = s$. Por ejemplo, la solución de $s(t) = 1$ es $t(1) \approx 0.892667771\dots$. Así podemos recorrer la espiral, por ejemplo, a intervalos de longitud 1, como muestra la figura de la izquierda. Por ejemplo, en la figura vemos que la primera media vuelta de la espiral mide aproximadamente 6 unidades y, en efecto, $s(\pi) \approx 6.11$. ■



Ejemplo: La hélice En algunos casos, las parametrizaciones por el arco son fáciles de calcular. Es del caso de la hélice (mostrada en la figura de la derecha):

$$\vec{r}(t) = (a \cos t, a \sin t, bt),$$

que gira alrededor del eje Z con radio $a > 0$ de modo que cada vez que da una vuelta completa ha ascendido $2\pi b$ unidades (o descendido, si $b < 0$). Entonces

$$\vec{v} = (-a \sin t, a \cos t, b), \quad v = \sqrt{a^2 + b^2}.$$

Por consiguiente, llamando $c = \sqrt{a^2 + b^2}$, tenemos que

$$s(t) = \int_0^t cu du = ct,$$

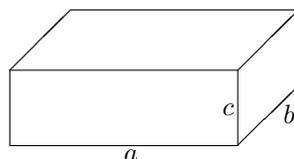
luego el cambio de parámetro que parametriza por el arco es $t = s/c$ y la reparametrización es

$$\vec{r}(s) = (a \cos(s/c), a \sin(s/c), bs/c). \quad \blacksquare$$

2.6 Cálculo de volúmenes

Veamos ahora que el cálculo integral que hemos expuesto también nos permite calcular algunos volúmenes. En el apéndice A de [ITAn] definimos la medida de Jordan en \mathbb{R}^2 , que formaliza la noción intuitiva de área, y toda la construcción se puede generalizar trivialmente para definir la medida de Jordan en \mathbb{R}^3 o, de hecho, en \mathbb{R}^n para un n arbitrario. La única parte que no es trivial es la prueba de que la medida de Jordan es invariante por isometrías. Ahora bien, en este libro no vamos a usar la noción de volumen más que en esta sección como mera ilustración de las posibilidades que ofrece el cálculo integral, así que aceptaremos sin más precisiones que es posible hablar del volumen de cualquier cuerpo tridimensional “razonable” de modo que se cumplen las propiedades obvias (como que si un cuerpo está contenida en otro, entonces su volumen es menor o igual).

El mismo razonamiento empleado en la sección A.1 de [ITAn] para justificar que el área de un rectángulo de lados a y b es igual (o debe definirse como igual) a ab prueba que el volumen de un prisma recto⁷ de base rectangular de aristas a, b, c debe ser igual a abc :



Se razona primero para el caso en que a, b, c son números racionales y en el caso general se prueba que todo prisma contiene y está contenido en prismas de aristas racionales con volúmenes tan próximos como se quiera a abc , por lo que éste es el único valor aceptable para su volumen.

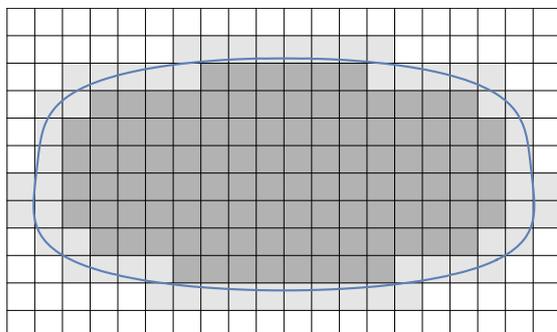
Un poco más en general: el volumen de un prisma recto generalizado, es decir, una figura limitada por dos bases de forma arbitraria situadas en planos paralelos y por una superficie lateral perpendicular a dichos planos, es el producto del área de la base por la altura Sh :



Este hecho admite una justificación sencilla: por el teorema [ITAn A.8], si B es la base del prisma generalizado P , existen figuras elementales $A_0 \subset B \subset A_1$ (que por [ITAn A.4] se pueden expresar como unión de un número finito de

⁷Un *prisma recto* es una figura formada limitada por dos polígonos paralelos (sus bases) unidos por rectángulos perpendiculares a los planos que los contienen.

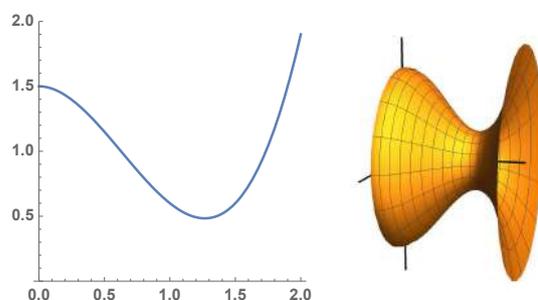
rectángulos disjuntos) cuya diferencia de áreas $m(A_1) - m(A_0)$ es tan pequeña como se quiera, luego ambas se aproximan a $S = m(B)$ tanto como se quiera:



Los prismas rectos de altura h construidos sobre los rectángulos que forman A_0 y A_1 suman un volumen igual a $m(A_0)h$ y $m(A_1)h$, respectivamente, de modo que uno está contenido y el otro contiene al prisma generalizado P y dichos volúmenes se aproximan tanto cuanto queramos a Sh , por lo que éste es necesariamente el volumen de P .

Estos hechos básicos sobre el volumen son suficientes para las aplicaciones de la integral que vamos a mostrar a continuación.

Volúmenes de sólidos de revolución Sea $f : [a, b] \rightarrow [0, +\infty[$ una función continua y vamos a calcular el volumen del sólido limitado por la superficie que resulta de hacer girar la curva alrededor del eje de abscisas:

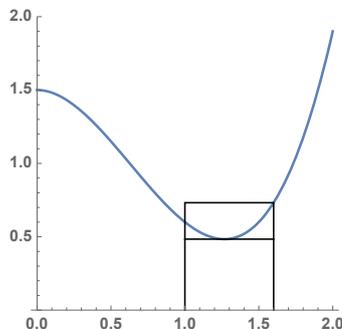


Para ello llamamos $V(x)$ a la función que da el volumen del sólido que resulta de girar la gráfica de la función restringida al intervalo $[a, x]$. En estos términos, queremos calcular $V(b)$. Notemos que V es una función creciente.

Fijemos $a < x < b$ y sea $h \neq 0$ suficientemente pequeño como para que $x + h$ esté en $[a, b]$. Por el teorema de Weierstrass [ITAn 3.31] existen puntos x_- y x_+ en el intervalo⁸ $[x, x + h]$ donde f toma su valor mínimo $m = f(x_-)$ y su valor máximo $M = f(x_+)$. Entonces $|V(x + h) - V(x)|$ es el volumen del sólido que

⁸Si $h < 0$ entendemos que $[x, x + h] = [x + h, x]$.

resulta de girar la gráfica de f restringida al intervalo $[x, x+h]$, el cual contiene un cilindro de radio m y altura $|h|$ y está contenido en un cilindro de radio M y altura $|h|$:



Por las observaciones iniciales sobre prismas generalizados, sabemos que el volumen de un cilindro de radio r y altura $|h|$ es $\pi r^2|h|$, luego podemos afirmar que

$$\pi m^2|h| \leq |V(x+h) - V(x)| \leq \pi M^2|h|,$$

luego

$$\pi m^2 \leq \frac{V(x+h) - V(x)}{h} \leq \pi M^2,$$

donde hemos quitado el valor absoluto porque el signo del numerador es el mismo que el del denominador, luego la fracción es positiva. A su vez,

$$\pi m^2 - \pi f(x)^2 \leq \frac{V(x+h) - V(x)}{h} - \pi f(x)^2 \leq \pi M^2 - \pi f(x)^2.$$

Ahora aplicamos el teorema de Heine-Cantor [ITAn A.14] a la función πf^2 , según el cual, dado $\epsilon > 0$, existe un $\delta > 0$ tal que si $u, v \in [a, b]$ cumplen $|u - v| < \delta$, entonces $|\pi f(u)^2 - \pi f(v)^2| < \epsilon$. Así, si restringimos el razonamiento precedente a valores $|h| < \delta$ y tomamos como u, v los puntos x y x_- , por una parte, y x y x_+ por otra, concluimos que $\pi f(x)^2 - \pi m^2 < \epsilon$ y $\pi M^2 - \pi f(x)^2 < \epsilon$, luego

$$-\epsilon < \frac{V(x+h) - V(x)}{h} - \pi f(x)^2 < \epsilon,$$

luego

$$\left| \frac{V(x+h) - V(x)}{h} - \pi f(x)^2 \right| < \epsilon,$$

y esto prueba que existe

$$V'(x) = \lim_{h \rightarrow 0} \frac{V(x+h) - V(x)}{h} = \pi f(x)^2.$$

Dejamos a cargo del lector la prueba de que V es continua en a y en b , con lo que resulta ser una primitiva en $[a, b]$ de la función $\pi f(x)^2$, luego el volumen $V = V(b) - V(a)$ es igual a:

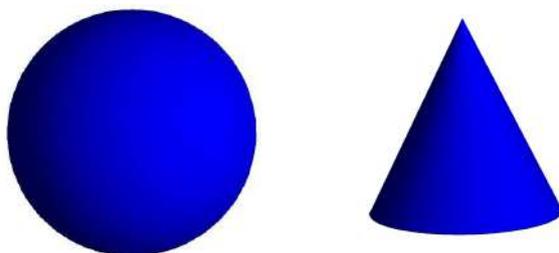
$$V = \pi \int_a^b f(x)^2 dx. \quad (2.4)$$

El volumen de una esfera Una esfera de radio r resulta del giro de la semicircunferencia $f(x) = \sqrt{r^2 - x^2}$, luego su volumen es

$$V = \pi \int_{-r}^r (r^2 - x^2) dx = \pi \left[r^2 x - \frac{x^3}{3} \right]_{-r}^r = \frac{4}{3} \pi r^3.$$

El volumen de un cono Un cono (recto) de radio r y altura h se obtiene del giro de la recta $f(x) = rx/h$ en $[0, h]$, luego su volumen es:

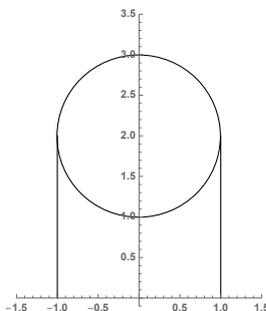
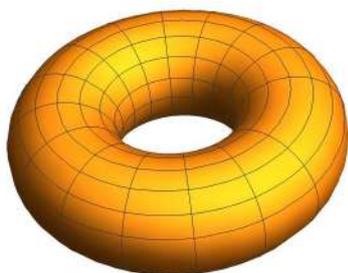
$$V = \pi \int_0^h \frac{r^2}{h^2} x^2 dx = \pi \left[\frac{r^2}{3h^2} x^3 \right]_0^h = \frac{1}{3} \pi r^2 h.$$



El volumen de un toro Un *toro* es el sólido que resulta del giro de una circunferencia alrededor de un eje exterior a ella. Esto no se corresponde exactamente con el tipo de sólidos de revolución que estamos considerando, pero podemos calcular su volumen como el volumen del sólido que resulta del giro de la semicircunferencia $f_+(x) = R + \sqrt{r^2 - x^2}$ menos el volumen del sólido que resulta del giro de la semicircunferencia opuesta: $f_-(x) = R - \sqrt{r^2 - x^2}$. Así, el volumen de un toro de radios $r < R$ viene dado por

$$V = \pi \int_{-r}^r (f_+^2(x) - f_-^2(x)) dx = 2\pi R \int_{-r}^r \sqrt{r^2 - x^2} dx = 2\pi R \frac{\pi r^2}{2},$$

pues la integral es la que da el área del semicírculo de radio r .



En definitiva, el volumen del toro es $V = 2\pi^2 Rr^2$, que puede interpretarse como el producto del área del círculo menor, πr^2 , por la longitud de la circunferencia de radio R , que es $2\pi R$. Podríamos decir que si tomamos un cilindro de radio r y altura $2\pi R$ y lo doblamos para formar un toro, su volumen no se altera.

En las secciones correspondientes del apéndice B calculamos el volumen de los sólidos de revolución determinados por la cicloide, la tractriz y la lemniscata de Bernoulli. Para el caso de la cardioide y la nefroide necesitamos las observaciones del apartado siguiente:

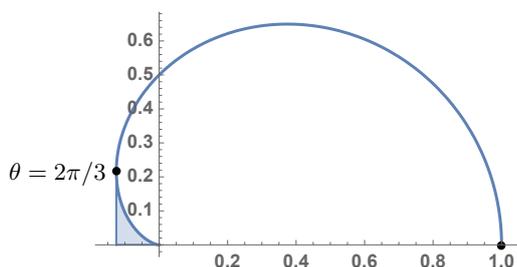
Sólidos determinados por curvas paramétricas Supongamos ahora que la curva que determina el sólido de revolución es una curva paramétrica, digamos $r(t) = (x(t), y(t))$, con $y(t) \geq 0$ en un intervalo $[a, b]$. Como en el caso del área, podemos razonar que el volumen del sólido de revolución que determina vendrá dado por

$$V = \pm\pi \int_a^b y^2(t)x'(t) dt. \quad (2.5)$$

En efecto, consideremos, por ejemplo el caso de la cardioide, para la cual

$$x = \rho \cos \theta, \quad y = \rho \operatorname{sen} \theta,$$

donde a su vez $\rho(\theta) = \frac{a}{2}(1 + \cos \theta)$.



La cardioide entre $\theta = 0$ y $\theta = 2\pi/3$ es la gráfica de una función $y(x)$, y el volumen del sólido de revolución generado por ese tramo (que incluye la zona sombreada en la figura) se puede calcular como

$$-\pi \int_{x(0)}^{x(2\pi/3)} y^2(x) dx = -\pi \int_0^{2\pi/3} y^2(t) x'(t) dt,$$

donde el signo negativo compensa que los extremos de la primera integral están al revés. Pero, por otra parte, la integral

$$-\pi \int_{x(2\pi/3)}^{x(\pi)} y^2(x) dx = -\pi \int_{2\pi/3}^{\pi} y^2(t) x'(t) dt,$$

nos da el volumen generado por la zona sombreada con signo negativo, porque ahora los extremos de la primera integral están al derecho. Por consiguiente (2.5)

es exactamente el volumen del sólido de revolución limitado por la cardioide, y el razonamiento es aplicable a cualquier curva paramétrica, con la única salvedad de que el signo negativo es necesario si la curva se recorre en sentido antihorario, pero no si se recorre en sentido horario.

Pero en el caso de una curva dada en coordenadas polares con parámetro θ , como es el caso de la cardioide, la fórmula puede simplificarse mucho integrando por partes:

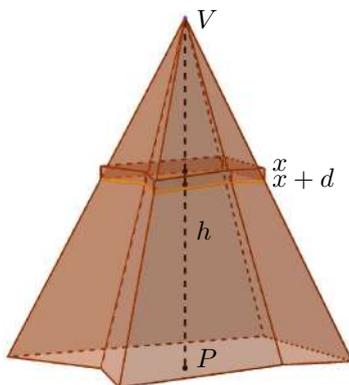
$$\begin{aligned} V &= -\pi \int_0^\pi \rho^2 \operatorname{sen}^2 \theta (\rho' \cos \theta - \rho \operatorname{sen} \theta) d\theta \\ &= -\pi \int_0^\pi (\rho^2 \rho' (1 - \cos^2 \theta) \cos \theta - \rho^3 \operatorname{sen}^3 \theta) d\theta \\ &= -\pi \int_0^\pi \rho^2 \rho' (\cos \theta - \cos^3 \theta) d\theta + \pi \int_0^\pi \rho^3 \operatorname{sen}^3 \theta d\theta \\ &= -\pi \left[(\cos \theta - \cos^3 \theta) \frac{\rho^3}{3} \right]_0^\pi + \pi \int_0^\pi \frac{\rho^3}{3} (-\operatorname{sen} \theta + 3 \cos^2 \theta \operatorname{sen} \theta) d\theta + \pi \int_0^\pi \rho^3 \operatorname{sen}^3 \theta d\theta \\ &= \pi \int_0^\pi \left(-\frac{\rho^2}{3} \operatorname{sen} \theta + \rho^3 (1 - \operatorname{sen}^2 \theta) \operatorname{sen} \theta + \rho^3 \operatorname{sen}^2 \theta \right) d\theta \end{aligned}$$

y en definitiva llegamos a que el volumen del sólido de revolución determinado por una ecuación $\rho = \rho(\theta)$ en coordenadas polares viene dado por

$$V = \frac{2\pi}{3} \int_0^\pi \rho^3 \operatorname{sen} \theta d\theta. \quad (2.6)$$

En la sección B.2 calculamos el volumen del sólido de revolución determinado por la cardioide usando esta fórmula. ■

Pirámides generalizadas La técnica que hemos usado para calcular el volumen de los sólidos de revolución se puede aplicar a cualquier pirámide generalizada limitada por una base plana arbitraria y por la superficie lateral que resulta de unir todos los puntos del borde de la base con un mismo vértice V . Aunque en realidad no es necesario, para que el argumento que vamos a dar funcione, debemos exigir que el pie P de la perpendicular por el vértice al plano de la base esté contenido en la base.



Es claro que si cortamos una pirámide generalizada con un plano paralelo a su base a una distancia x de su vértice la figura B_x que obtenemos es una versión a escala de la base, en proporción x/h , donde h es la altura de la pirámide. Si la base tiene área S , entonces el área de B_x es $S(x/h)^2$ (esto es esencialmente consecuencia del hecho de que una homotecia de razón r multiplica las áreas por r^2).

Llamemos $V(x)$ al volumen de la pirámide que resulta de cortar la pirámide dada por un plano paralelo a su base a una distancia x de su vértice. Entonces, si $d > 0$, la diferencia $V(x+d) - V(x)$ es el volumen de una “rodaja” de pirámide que contiene un prisma de base B_x y altura d y está contenido en un prisma de base B_{x+d} y altura d (es aquí donde necesitamos suponer que el pie P de la altura de la pirámide está contenido en la base, aunque insistimos en que la conclusión a la que vamos a llegar es cierta aunque esto no se cumpla). Por lo tanto,

$$S \frac{x^2}{h^2} d \leq V(x+d) - V(x) \leq S \frac{(x+d)^2}{h^2} d = S \frac{x^2}{h^2} d + S d^2 \frac{2x+d}{h^2}$$

luego

$$0 \leq \frac{V(x+d) - V(x)}{d} - \frac{S}{h^2} x^2 \leq S d \frac{2x+d}{h^2}.$$

Si $d < 0$ las desigualdades se invierten, pero en ambos casos se cumple que

$$\left| \frac{V(x+d) - V(x)}{d} - \frac{S}{h^2} x^2 \right| \leq S |d| \frac{2x+d}{h^2} \leq 2S |d|,$$

donde hemos acotado $x, x+d \leq h$. Es claro entonces que existe

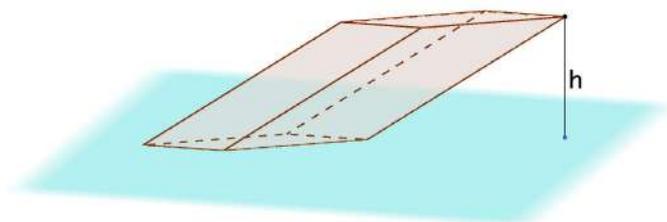
$$\frac{dV}{dx} = \lim_{d \rightarrow 0} \frac{V(x+d) - V(x)}{d} = \frac{S}{h^2} x^2.$$

También es fácil ver que V es continua en 0 y en h (en realidad en h no hay nada que comprobar, pues siempre podemos considerar una pirámide más alta, con lo que V está definida y es derivable en realidad en el intervalo $]0, +\infty[$). Así podemos concluir que el volumen de la pirámide de altura h es

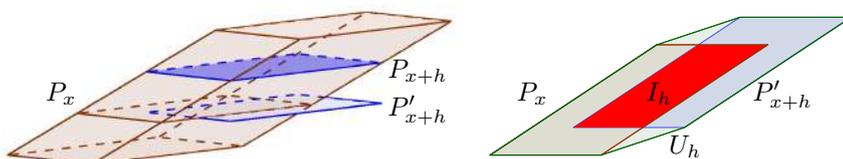
$$V = \int_0^h \frac{S}{h^2} x^2 dx = \frac{S}{h^2} \left[\frac{x^3}{3} \right]_0^h = \frac{1}{3} S h.$$

En otras palabras: el volumen de una pirámide generalizada es un tercio de la superficie de su base por su altura. El volumen del cono que habíamos calculado anteriormente es un caso particular de este hecho. ■

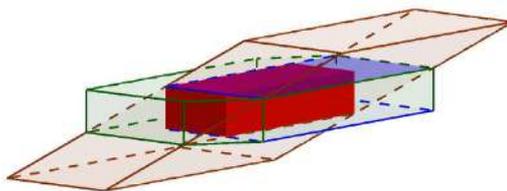
Paralelepípedos Vamos a probar ahora que el volumen de un paralelepípedo como el de la figura siguiente es igual a la superficie B de su base por su altura h , es decir, que al “inclinarse” un paralelepípedo no altera su volumen mientras que mantenga su altura. Para probarlo necesitamos refinar el argumento empleado con las pirámides rectas, pues las “rodajas” en forma de prisma análogas a las que consideramos en ese caso no cumplen que una esté contenida en la rodaja de paralelepípedo correspondiente y otra lo contenga.



Sea $V(x)$ el volumen de la porción del paralelepípedo situado por debajo del plano Π_x paralelo a la base situado a una distancia x de ésta. La intersección del paralelepípedo con Π_x es un paralelogramo P_x igual a la base. Similarmente, el plano Π_{x+h} determina un paralelogramo P_{x+h} que podemos proyectar verticalmente a Π_x para obtener otro paralelogramo P'_{x+h} .



La figura de la derecha corresponde al plano Π_x . Los paralelogramos P_x y P'_{x+h} son ambos iguales a la base del paralelepípedo, y es claro que a medida que h tiende a 0 el paralelepípedo P'_{x+h} se va acercando a P_x hasta superponerse con él. Llamamos I_h a la intersección de ambos y U_h al polígono que muestra la figura, de modo que el prisma recto de altura h y base I_h está contenido en el paralelepípedo y el de altura h y base U_h contiene a la sección del paralelepípedo comprendida entre los planos Π_x y Π_{x+h} , tal y como muestra la figura siguiente:



Puesto que sabemos calcular el área de un prisma recto, podemos afirmar que

$$i_h h \leq V(x+h) - V(x) \leq u_h h,$$

donde i_h es el área de I_h y u_h es el área de U_h . Equivalentemente:

$$i_h \leq \frac{V(x+h) - V(x)}{h} \leq u_h$$

Ahora bien, el hecho de que P'_{x+h} se mueva hasta superponerse con P_x cuando h tiende a 0 se traduce claramente en que $\lim_{h \rightarrow 0} i_h = \lim_{h \rightarrow 0} u_h = B$ es el área de la base del paralelepípedo.

De hecho, podríamos calcular explícitamente i_h y u_h en función del ángulo que las aristas laterales del paralelepípedo forman con el plano de su base y constatar que ambos tienden a B , pero con tales cálculos sólo estaríamos constataando lo que ya es obvio sin necesidad de ellos. La conclusión es que

$$\frac{dV}{dx} = B,$$

luego, teniendo en cuenta además que $V(0) = 0$, concluimos que $V(x) = Bx$ y, en particular, el volumen del paralelepípedo completo es $V(h) = Bh$, como queríamos probar. ■

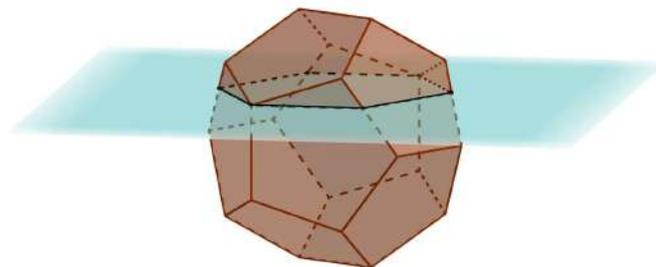
Más en general, puede demostrarse que si tomamos cualquier sólido “razonable” y consideramos sus secciones por un haz de planos paralelos, de modo que $B(x)$ es el área de la sección determinada por el plano situado a altura x y $V(x)$ es el volumen de la parte del sólido situada bajo dicho plano, entonces

$$\frac{dV}{dx} = B(x),$$

por lo que el volumen del sólido es

$$V = \int_a^b B(x) dx,$$

donde a y b son la menor y la mayor altura posible de un plano que corte al sólido.



Cuando la geometría del sólido no es excesivamente complicada, esto puede probarse modificando el argumento con el que lo hemos probado para el caso de los paralelepípedos. Por ejemplo, no es difícil probar que esto es cierto para cilindros y pirámides con base arbitraria, no necesariamente rectos, pero una prueba general requiere prescindir de consideraciones particulares, y entonces la dificultad estriba en precisar qué entendemos por “sólido razonable”, pues no es posible asignar un volumen a cualquier región del espacio, aunque esté acotada. Esto requiere introducir la llamada “medida de Lebesgue”, y el resultado es un caso particular del llamado “teorema de Fubini”. La demostración no es complicada, pero se enmarca en el contexto de la teoría de la medida abstracta, que dista mucho del cálculo integral que hemos visto aquí. En particular, de

este resultado (aplicable, digamos, a figuras “razonables”, que podamos “esculpir” fidedignamente) se deduce el principio siguiente, que ya hemos discutido en la introducción:

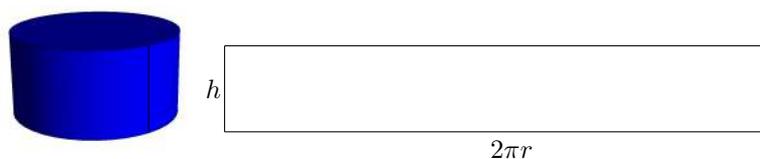
Principio de Cavalieri *Si dos figuras están limitadas entre dos mismos planos paralelos, y cualquier plano paralelo intermedio determina en ambas secciones con la misma área, entonces ambas tienen el mismo volumen.*

La razón es que la función $B(x)$ es la misma para ambas figuras, por lo que ambos volúmenes coinciden con la integral de $B(x)$ entre las alturas de los planos paralelos que limitan ambas figuras.

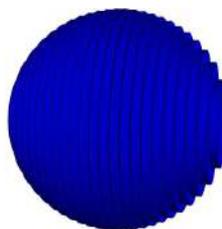
2.7 Áreas de sólidos de revolución

En la sección anterior hemos señalado que definir el volumen de una figura tridimensional “razonable” no requiere más que generalizar de forma obvia la definición de medida de Jordan que presentamos en el apéndice A de [ITAn] para el caso bidimensional. Sin embargo, definir el área de un sólido resulta ser un problema mucho más delicado. Así, antes de preguntarnos, por ejemplo, cuál es el área de una esfera, tendríamos que preguntarnos a qué llamamos área de una esfera. Una vez más, como sólo vamos a hablar de áreas de sólidos como una aplicación aislada del cálculo integral, no trataremos de dar una definición general de “área” de un sólido, sino que aceptaremos que hay algo llamado área que está esperando a ser medido y veremos que cierta fórmula integral proporciona un valor que es razonable identificar con el área de un sólido de revolución.

En primer lugar, no hay duda de cuál tiene que ser el área de un cilindro de radio r y altura h : si hacemos un corte en la superficie lateral de un cilindro perpendicularmente a sus bases obtenemos una cinta que podemos desplegar hasta transformarla en un rectángulo de base $2\pi r$ y altura h , luego su área es $2\pi rh$. Ésta tiene que ser, pues, el área (de la superficie lateral) de un cilindro.

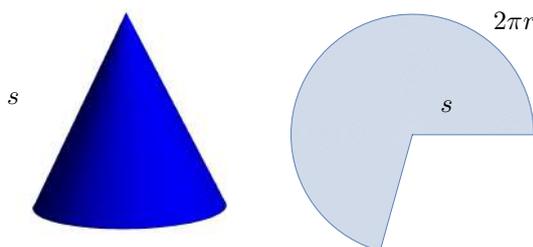


A partir de aquí, podríamos tratar de calcular el área de una superficie de revolución descomponiéndola en rodajas y aproximando la superficie de cada rodaja por la de un cilindro. La figura siguiente muestra el caso de una esfera:

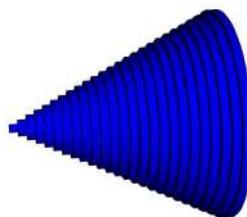


En la sección anterior hemos demostrado que la suma de los volúmenes de las rodajas cilíndricas se aproxima al volumen de la esfera, y que la aproximación es mejor cuanto mayor es el número de rodajas. Ello podría llevarnos a conjeturar que la suma de las áreas de las rodajas tiende al área de la esfera. Sin embargo, no es así. Para probar que el “método” no sirve vamos a usarlo para calcular (mal) el área de otra figura que, por otra parte, podemos calcular fácilmente de forma alternativa.

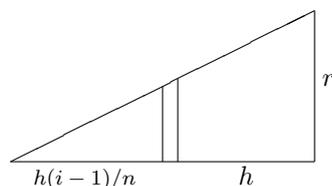
Si cortamos un cono de radio r y altura h por una de sus generatrices obtenemos un sector circular de radio igual a la generatriz del cono, es decir, $s = \sqrt{r^2 + h^2}$, y cuya longitud de arco es $2\pi r$, luego su amplitud es $2\pi r / 2\pi s = r/s$, luego su área es $\pi s^2 r / s = \pi r s = \pi r \sqrt{r^2 + h^2}$.



Ahora trataremos de llegar a esta misma conclusión aproximando el cono mediante rodajas cilíndricas:



Para ello dividimos la altura h en n partes iguales, y en cada una de ellas, digamos en el intervalo $[h(i-1)/n, hi/n]$, consideramos el cilindro de altura h/n y radio $h(i-1)/n r/h = r(i-1)/n$:



Así, el área y el volumen de la rodaja i -ésima es

$$A_i = 2\pi r h (i-1)/n^2, \quad V_i = \pi r^2 h (i-1)^2/n^3,$$

y las aproximaciones al área y el volumen del cono mediante las rodajas son, respectivamente:

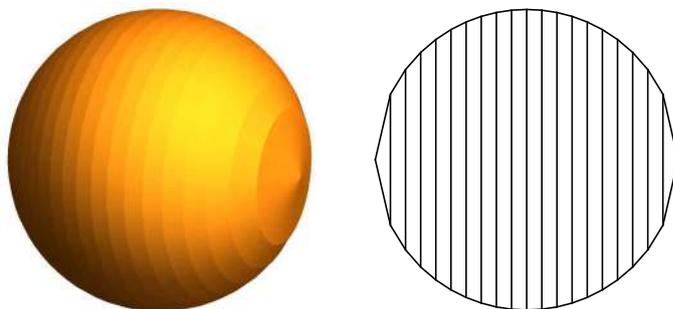
$$\sum_{i=1}^n 2\pi r h \frac{i-1}{n^2} = \frac{2\pi r h (n-1)n}{n^2} = \pi r h \frac{n-1}{n} \rightarrow \pi r h,$$

$$\sum_{i=1}^n \pi r^2 h \frac{(i-1)^2}{n^3} = \frac{\pi r^2 h (n-1)n(2n-1)}{n^3} = \frac{\pi r^2 h}{6} \frac{n-1}{n} \frac{2n-1}{n} \rightarrow \frac{1}{3} \pi r^2 h.$$

Vemos que las aproximaciones del volumen tienden al volumen del cono, pero las “aproximaciones” del área no tienden al área del cono que hemos calculado previamente.

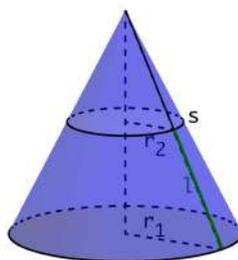
La diferencia es que cuando hemos argumentado que las sumas de volúmenes de los cilindros tienden al volumen del cono (o de cualquier sólido de revolución) hemos visto que podíamos considerar cilindros contenidos en las rodajas y cilindros que contienen las rodajas, y que en ambos casos el límite de los volúmenes es el mismo, por lo que dicho límite es necesariamente el volumen del sólido que queremos calcular. En cambio, en el caso del cono, es fácil ver que las superficies de los cilindros son menores o iguales que las áreas de las rodajas, pero si consideramos cilindros que contengan a las rodajas no podemos probar que las áreas de los cilindros sean mayores o iguales que las de las rodajas, así que, aunque el área calculada con los cilindros mayores coincide con la calculada con los cilindros menores, no tenemos nada que nos asegure que dicha área tiene que ser la del cono, y ciertamente no lo es.

Consideremos ahora la figura siguiente:



Parece una esfera cortada en rodajas, pero no es eso exactamente, pues las rodajas no son “aros” de esfera, sino de cono. La figura de la derecha muestra una sección de la figura. Aumentando el número de divisiones podríamos hacer que el contorno de la figura de la derecha fuera indistinguible de una auténtica circunferencia, y entonces la figura de la izquierda sería indistinguible de una esfera, por lo que es razonable considerar que las áreas de estas figuras formadas por troncos de cono pueden aproximar el área real de una esfera —o, más en general, de cualquier sólido de revolución— con tanta precisión como queramos. Partiendo de esta idea, vamos a obtener una fórmula para el área de un sólido de revolución.

Para ello necesitamos en primer lugar una fórmula para el área de un tronco de cono.



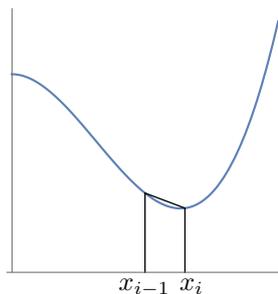
Podemos expresarla como la diferencia entre las áreas de dos conos de generatrices s y $s - l$ y radios r_1 y r_2 , respectivamente, tenemos que

$$\frac{r_2}{r_1} = \frac{s - l}{s},$$

luego $r_2 s = r_1 s - r_1 l$, luego $s = (r_1 l) / (r_1 - r_2)$. Por consiguiente, el área del tronco de cono es

$$\begin{aligned} S &= \pi r_1 s - \pi r_2 (s - l) = \pi r_1 \frac{r_1 l}{r_1 - r_2} - \pi r_2 \left(\frac{r_1 l}{r_1 - r_2} - l \right) \\ &= \frac{\pi r_1^2 l}{r_1 - r_2} - \frac{\pi r_2^2 l}{r_1 - r_2} = \pi \frac{(r_1 + r_2)(r_1 - r_2) l}{r_1 - r_2} = \pi (r_1 + r_2) l = 2\pi \frac{r_1 + r_2}{2} l. \end{aligned}$$

Ahora consideremos una función $f : [a, b] \rightarrow [0, +\infty[$ continua, derivable en $]a, b[$ y cuya derivada admita una extensión continua $f' : [a, b] \rightarrow \mathbb{R}$.



Para calcular el área del sólido que resulta de girar la gráfica de f alrededor del eje de abscisas, partimos $[a, b]$ en n intervalos $a = x_0 < x_1 < \dots < x_n = b$ de longitud $\Delta x_i = x_i - x_{i-1}$, de modo que el área del tronco de cono determinado por los radios $f(x_{i-1})$ y $f(x_i)$ y la altura Δx_i es

$$S_i = 2\pi \frac{f(x_{i-1}) + f(x_i)}{2} \sqrt{\Delta x_i^2 + (f(x_i) - f(x_{i-1}))^2},$$

y la suma de todas estas áreas es:

$$S = 2\pi \sum_{i=1}^n 2\pi \frac{f(x_{i-1}) + f(x_i)}{2} \sqrt{\Delta x_i^2 + (f(x_i) - f(x_{i-1}))^2}.$$

Por el teorema de los valores intermedios [ITAn 3.24] y el teorema del valor medio, existen $\eta_i, \xi_i \in [x_{i-1}, x_i]$ tales que

$$\frac{f(x_{i-1}) + f(x_i)}{2} = f(\eta_i), \quad f(x_i) - f(x_{i-1}) = f'(\xi_i)\Delta x_i.$$

Por consiguiente, $S_i = 2\pi f(\eta_i)\sqrt{1 + f'(\xi_i)^2}\Delta x_i$. Ahora aplicamos el teorema de Heine-Cantor [ITAn A.14] a la función $g : [a, b] \times [a, b] \rightarrow \mathbb{R}$ dada por $g(x, y) = 2\pi f(x)\sqrt{1 + f'(y)^2}$:

Dado $\epsilon > 0$, existe un $\delta^* > 0$ tal que si $\|(x, y) - (x', y')\| < \delta^*$, entonces $|g(x, y) - g(x', y')| < \epsilon/2(b-a)$.

Así, si tomamos una partición que cumpla $\Delta x_i < \delta = \delta^*/\sqrt{2}$ para todo i , entonces $\|(\eta_i, \xi_i) - (x_i, x_i)\| < \delta$, luego

$$|2\pi f(\eta_i)\sqrt{1 + f'(\xi_i)^2} - 2\pi f(x_i)\sqrt{1 + f'(x_i)^2}| < \frac{\epsilon}{2(b-a)}.$$

Por lo tanto,

$$\begin{aligned} & \left| S - 2\pi \sum_{i=1}^n f(x_i)\sqrt{1 + f'(x_i)^2} \Delta x_i \right| = \\ & \left| \sum_{i=1}^n 2\pi f(\eta_i)\sqrt{1 + f'(\xi_i)^2} \Delta x_i - f(x_i)\sqrt{1 + f'(x_i)^2} \Delta x_i \right| \leq \sum_{i=1}^n \frac{\epsilon}{2(b-a)} \Delta x_i = \frac{\epsilon}{2}. \end{aligned}$$

Por otra parte, el teorema 2.5 nos da que podemos elegir δ suficientemente pequeño como para que, además, si $\Delta x_i < \delta$, entonces

$$\left| 2\pi \sum_{i=1}^n f(x_i)\sqrt{1 + f'(x_i)^2} - 2\pi \int_a^b f(x)\sqrt{1 + f'(x)^2} dx \right| < \frac{\epsilon}{2}.$$

Combinando ambas desigualdades tenemos que, para todo $\epsilon > 0$, existe un $\delta > 0$ tal que si partimos el intervalo $[a, b]$ en subintervalos de longitud $\Delta x_i < \delta$, entonces

$$\left| S - 2\pi \int_a^b f(x)\sqrt{1 + f'(x)^2} dx \right| < \epsilon.$$

Con esto podemos concluir que el área del sólido de revolución definido por una función continua $f : [a, b] \rightarrow [0, +\infty[$, derivable en $]a, b[$ cuya derivada se extienda a una función continua en $[a, b]$ viene dada por

$$S = 2\pi \int_a^b f(x) \sqrt{1 + f'(x)^2} dx, \quad (2.7)$$

en el sentido de que esta integral es el límite al que se aproximan las áreas de las aproximaciones del sólido formadas por troncos de cono cuando las alturas de éstos tienden a 0.

El área de una esfera En el caso de una esfera, $f(x) = \sqrt{r^2 - x^2}$, y la fórmula anterior nos da

$$S = 2\pi \int_{-r}^r \sqrt{r^2 - x^2} \sqrt{1 + \frac{x^2}{r^2 - x^2}} dx = 2\pi \int_{-r}^r r dx = 4\pi r^2.$$

El área de un toro El área de un toro se puede calcular como la suma de las áreas de los sólidos de revolución generados por las funciones $R \pm \sqrt{r^2 - x^2}$:

$$\begin{aligned} & 2\pi \int_{-r}^r (R + \sqrt{r^2 - x^2}) \sqrt{1 + \frac{x^2}{r^2 - x^2}} dx + 2\pi \int_{-r}^r (R - \sqrt{r^2 - x^2}) \sqrt{1 + \frac{x^2}{r^2 - x^2}} dx \\ & 4\pi \int_{-r}^r R \frac{r}{\sqrt{r^2 - x^2}} dx = 4\pi Rr \int_{-r}^r \frac{1/r}{\sqrt{1 - (x/r)^2}} dx = 4\pi Rr [\arcsen(x/r)]_{-r}^r, \end{aligned}$$

luego el área es $S = 4\pi^2 Rr$.

El cuerno de Gabriel Un ejemplo muy sencillo, pero a la vez muy interesante, lo proporciona el sólido de revolución generado por la curva $f(x) = 1/x$ desde $x = 1$:



Su peculiaridad radica en que, como ya indicamos en la introducción, su volumen es finito, pero su área es finita. En efecto, el volumen del “cuerno” es

$$V = \int_1^{+\infty} \frac{\pi}{x^2} dx = \lim_{x_0 \rightarrow +\infty} \int_1^{x_0} \frac{\pi}{x^2} dx = \lim_{x_0 \rightarrow +\infty} \pi \left[\frac{-1}{x} \right]_1^{x_0} = \lim_{x_0 \rightarrow +\infty} \pi - \frac{\pi}{x_0} = \pi.$$

En cambio, su superficie es

$$\begin{aligned} S &= \int_1^{+\infty} \frac{2\pi}{x} \sqrt{1 + \frac{1}{x^4}} dx = \lim_{x_0 \rightarrow +\infty} \int_1^{x_0} \frac{2\pi}{x} \sqrt{1 + \frac{1}{x^4}} dx \geq \lim_{x_0 \rightarrow +\infty} \int_1^{x_0} \frac{2\pi}{x} dx \\ &= \lim_{x_0 \rightarrow +\infty} 2\pi \log x_0 = +\infty, \end{aligned}$$

luego el área es, ciertamente, infinita.

Por otra parte, es imposible que un sólido de revolución tenga área finita y volumen infinito. En efecto, observemos en primer lugar que si una función $f : [a, +\infty[\rightarrow [0, +\infty[$ determina un sólido de revolución con área finita S , entonces está acotada, pues

$$f(t)^2 - f(a)^2 = \int_a^t 2f(x)f'(x) dx \leq \int_a^t 2f(x)|f'(x)| dx \leq \int_a^t 2f(x)\sqrt{1+f'(x)^2} dx = \frac{1}{\pi}S(t) \leq \frac{1}{\pi}S.$$

Por lo tanto, $f(t) \leq \sqrt{f(a)^2 + S/\pi} = M$. Por consiguiente, el volumen hasta $x = t$ cumple

$$V(t) = \pi \int_a^t f(x)f(x) dx \leq M\pi \int_a^t f(x) dx \leq M\pi \int_a^t f(x)\sqrt{1+f'(x)^2} dx = \frac{M}{2}S(t) \leq \frac{M}{2}S,$$

luego, teniendo en cuenta que la función $V(t)$ es creciente, tiene que existir el límite

$$V = \int_a^{+\infty} \pi f^2(t) dt = \lim_{t \rightarrow +\infty} V(t),$$

es decir, que el volumen del sólido tiene que ser finito. ■

Superficies determinadas por curvas paramétricas Dada una curva paramétrica $r(t) = (x(t), y(t))$ con $y(t) \geq 0$ que no se corte a sí misma, en un intervalo $[a, b]$ en el que $x(t)$ sea creciente o decreciente, podemos hacer el cambio de variable

$$\begin{aligned} 2\pi \int_{x(a)}^{x(b)} y(x)^2 \sqrt{1+y'(x)^2} dx &= 2\pi \int_a^b y(t) \sqrt{1+y'(x(t))^2} x'(t) dt \\ &= \pm 2\pi \int_a^b y(t) \sqrt{1+y'(x(t))^2} |x'(t)| dt = \\ &= \pm 2\pi \int_a^b y(t) \sqrt{x'(t)^2 + (y'(x(t))x'(t))^2} dt = \pm 2\pi \int_a^b y(t) \sqrt{x'(t)^2 + y'(t)^2} dt, \end{aligned}$$

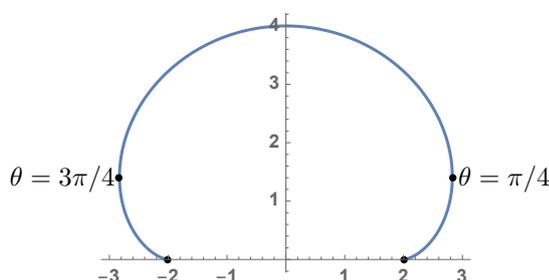
donde el signo \pm es el signo de $x'(t)$. Si es positivo, entonces $x(a) < x(b)$, luego la integral inicial es el área de la superficie de revolución generada con el giro del segmento de curva correspondiente a $a \leq t \leq b$, mientras que si es negativo entonces $x(a) > x(b)$ y la primera integral es negativa, luego en ambos casos la integral

$$2\pi \int_a^b y(t) \sqrt{x'(t)^2 + y'(t)^2} dt$$

es positiva (cosa que, por otro lado, es evidente, porque el integrando es positivo) y representa el área del fragmento de curva considerado. Si descomponemos el dominio completo $[a, b]$ de la curva en intervalos donde $x(t)$ sea creciente o decreciente, y sumar las integrales correspondientes, concluimos que el área de la superficie de revolución determinada por la curva es

$$A = 2\pi \int_a^b y(t) \sqrt{x'(t)^2 + y'(t)^2} dt \quad (2.8)$$

Por ejemplo, en el caso de la nefroide, cuando integramos en $[0, \pi/4]$ obtenemos el área generada por el primer segmento señalado en la figura, cuando integramos en $[\pi/4, 3\pi/4]$ obtenemos el área generada por el segundo y cuando integramos en $[3\pi/4, \pi]$ obtenemos el área generada por el tercero.



En el apéndice B calculamos con esta fórmula el área de los sólidos de revolución generados por la cicloide, la nefroide y la tractriz. Para la cardioide y la lemniscata conviene deducir una fórmula para curvas en polares:

$$A = 2\pi \int_a^b \rho \operatorname{sen} \theta \sqrt{(\rho' \cos \theta - \rho \operatorname{sen} \theta)^2 + (\rho' \operatorname{sen} \theta + \rho \cos \theta)^2} d\theta,$$

que se simplifica hasta

$$A = 2\pi \int_a^b \rho \operatorname{sen} \theta \sqrt{\rho^2 + \rho'^2} d\theta. \quad (2.9)$$

2.8 Ecuaciones diferenciales II

En la sección 1.7 hemos resuelto algunas ecuaciones diferenciales lineales por el método rudimentario de encontrar a ojo unas soluciones y justificar por un teorema de unicidad que no hay más. Aquí vamos a proporcionar una técnica para resolver una amplia familia de ecuaciones diferenciales de primer orden. Observemos en primer lugar que la ecuación más simple posible:

$$\frac{dy}{dx} = g(x), \quad y(x_0) = y_0$$

tiene por solución

$$y(x) = y_0 + \int_{x_0}^x g(t) dt,$$

pues la integral es de la forma $F(x) - F(x_0)$, para cierta primitiva F del integrando, luego $y_0 + F(x) - F(x_0)$ es la primitiva del integrando que vale y_0 en x_0 .

Más en general, vamos a resolver las ecuaciones diferenciales de la forma

$$f(y) \frac{dy}{dx} = g(x),$$

para ciertas funciones f y g . Se llaman *ecuaciones diferenciales de variables separables* porque pueden expresarse equivalentemente en la forma

$$f(y) dy = g(x) dx,$$

con todas las y 's en un miembro y todas las x 's en el otro.⁹

Una función $y(x)$, para ser solución de una ecuación de este tipo, debe cumplir que

$$f(y(x)) \frac{dy}{dx} = g(x).$$

Si f y g admiten primitivas F y G , respectivamente, entonces la función

$$f(y(x)) \frac{dy}{dx} - g(x)$$

es la derivada de $F(y(x)) - G(x)$, luego, en virtud del teorema 1.15, afirmar que la función anterior es idénticamente nula equivale a afirmar que $F(y(x)) - G(x)$ es constante. Así pues:

Si $f, g :]a, b[\rightarrow \mathbb{R}$ son funciones con primitivas F y G , respectivamente, las funciones $y :]a, b[\rightarrow \mathbb{R}$ que cumplen la ecuación

$$f(y) \frac{dy}{dx} = g(x)$$

son las mismas que cumplen la ecuación

$$F(y(x)) = G(x) + c,$$

para una constante c arbitraria.

Transformar una ecuación diferencial en otra ecuación equivalente en la que no aparezcan derivadas se llama *integrar* la ecuación. Acabamos de ver cómo integrar cualquier ecuación diferencial con variables separables.

En la práctica, más que calcular aparte las primitivas F y G y recordar el resultado que hemos probado, es más conveniente observar que podemos llegar a la ecuación $F(y) = G(x) + c$ integrando formalmente la igualdad anterior:

$$\int f(y) dy = \int g(x) dx,$$

⁹Por ejemplo, la ecuación diferencial $y'(x) = xy(x) + 3x^2$ no es de variables separables.

interpretando y en el miembro izquierdo como la variable de la función f , pues la solución de esta ecuación es,¹⁰ ciertamente, $F(y) = G(x) + c$, que sabemos que es la integral correcta de la ecuación diferencial.

Vemos que las soluciones de una ecuación diferencial de variables separables dependen en general de una constante arbitraria que podemos determinar mediante una condición inicial $y(x_0) = y_0$.

Ejemplo: El problema de Beaume El problema de Beaume —que, como hemos indicado en la introducción, Beaume planteó a Descartes y fue resuelto por Leibniz— consiste en encontrar las curvas con subtangente de longitud constante c , lo cual, según (1.3) equivale a resolver la ecuación diferencial

$$\frac{dy}{dx} = \frac{y}{c}.$$

Separando las variables:

$$\frac{dy}{y} = \frac{1}{c} dx,$$

luego

$$\int \frac{dy}{y} = \frac{1}{c} \int dx,$$

luego $\log |y| = x/c + k_0$, luego $|y| = e^{k_0} e^{x/c}$ y, llamando $k = \pm e^{k_0}$, vemos que la solución es $y = ke^{x/c}$. ■

Ejemplo: El crecimiento logístico En 1798 Thomas Robert Malthus publicó *Un ensayo sobre el principio de la población*, en el que exponía su teoría de que las poblaciones (tanto animales, o vegetales, o humanas) tienden a mostrar un crecimiento en progresión geométrica en el tiempo, mientras que los recursos disponibles (principalmente las fuentes de alimentos) crecen en progresión aritmética, por lo que sólo es cuestión de tiempo que se agoten los recursos disponibles.

Un modelo matemático de la teoría de Malthus se obtiene de la hipótesis de que si la función $P(t)$ representa el número de individuos de una población en cada instante t (digamos con el tiempo medido en años), entonces la tasa de crecimiento de P es proporcional a la población, es decir, que se cumple la ecuación diferencial

$$\frac{dP}{dt} = kP,$$

donde k es una constante que depende esencialmente de la tasa de natalidad y la tasa de mortalidad de la población.

Se trata de la ecuación diferencial que hemos resuelto en el teorema 1.32. La solución es, pues, que la población en un instante t será $P = P_0 e^{kt}$. Vemos que, en efecto para $t = 0, 1, 2, \dots$ el número de habitantes forma una progresión

¹⁰Podríamos escribir $F(y) + c_1 = G(x) + c_2$, pero a continuación podríamos llamar $c = c_2 - c_1$ y tendríamos la expresión que hemos escrito.

geométrica de razón $e^k > 1$, y a partir de aquí Malthus deducía que era urgente fomentar toda clase de medidas que aumentaran la mortalidad de los pobres para evitar el agotamiento de los recursos naturales.

Sin embargo, en 1845 Pierre François Verhulst publicó un modelo alternativo de crecimiento de poblaciones basado en la *ecuación logística*

$$\frac{dP}{dt} = kP \left(1 - \frac{P}{K}\right)$$

que resulta de añadir a la ecuación de Malthus un factor en el que K representa la máxima población que podría sobrevivir con los recursos disponibles, de modo que P/K indica la distancia de la población actual a dicha población máxima, y así el factor añadido expresa que el aumento de la población se ralentiza a medida que la población se acerca a la capacidad máxima que permiten los recursos.

A la hora de resolver esta ecuación es más cómodo llamar $x = P/K$, con lo que dividiendo entre K la ecuación se transforma en

$$\frac{dx}{dt} = kx(1-x).$$

Consideramos valores $0 < x < 1$ (que corresponden a poblaciones “viables” $0 < P < K$). Separando variables:

$$\frac{dx}{x(1-x)} = k dt.$$

Para integrar el miembro izquierdo lo descomponemos en suma de dos fracciones:

$$\int \left(\frac{1}{x} + \frac{1}{1-x} \right) dx = \int k dt.$$

Por lo tanto:

$$\log x + \log(1-x) = kt + c.$$

o también

$$\log \frac{x}{1-x} = kt + c,$$

luego

$$\frac{x}{1-x} = e^{kt+c}.$$

Esta ecuación tiene solución:

$$x = \frac{e^{kt+c}}{1 + e^{kt+c}} = \frac{e^{kt}}{e^{kt} + e^{-c}}.$$

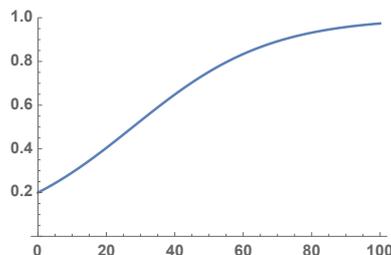
Si llamamos $C = e^{-c} > 0$, podemos expresarla en la forma

$$x(t) = \frac{e^{kt}}{e^{kt} + C} = \frac{1}{1 + Ce^{-kt}},$$

o también:

$$P(t) = \frac{K}{1 + Ce^{-kt}}.$$

La gráfica siguiente muestra la evolución de una población que es inicialmente $x(0) = 0.2$, es decir, un 20% de la población máxima, con $k = 0.05$:

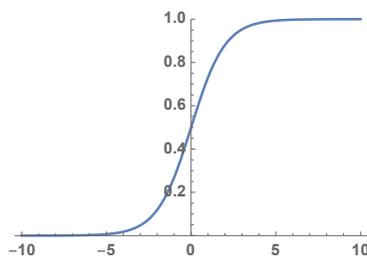


Es fácil ver en general que $\lim_{t \rightarrow +\infty} P(t) = K$, así como que la gráfica tiene un punto de inflexión en $t = \log C/k$, es decir, un punto donde pasa de ser convexa a ser cóncava.

En general, las soluciones de la ecuación logística se llaman *funciones logísticas*, entre las cuales la más simple es la que resulta de tomar todas las constantes iguales a 1:

$$f(x) = \frac{1}{1 + e^{-x}},$$

que tiene muchas aplicaciones porque pasa rápidamente de valer “casi 0” a valer “casi 1” (y esta “rapidez” se puede aumentar ajustando las constantes):



Rozamiento en el aire En el ejemplo siguiente al teorema 1.10 vimos que si lanzamos un objeto hacia arriba a una velocidad v , su altura en función del tiempo es $h(t) = vt - 4.9t^2$ m, la altura máxima a la que llega es $0.051v^2$, y es claro que el tiempo que tarda en llegar de nuevo al punto de partida es $v/4.9$ s. Por ejemplo, si la velocidad inicial es $v = 10$ m/s, el objeto llegará hasta los 5.1 m de altura y llegará al punto de partida al cabo de 2.04 s.

Ahora bien, esto es así si no tenemos en cuenta el rozamiento del aire, que frena el movimiento. En tal caso, ya no es cierto que la aceleración del objeto cumpla

$$\frac{dv}{dt} = -g \text{ m/s}^2,$$

donde $g = 9.8 \text{ m/s}^2$, sino que, mientras el objeto sube, el rozamiento del aire disminuye su velocidad proporcionalmente al cuadrado de su velocidad, es decir, de modo que

$$\frac{dv}{dt} = -g - kv^2 \text{ m/s}^2,$$

donde la constante k depende de muchos factores. Para el caso de una pelota de tenis podemos tomar

$$k = \frac{1}{2} \frac{1.2 \pi 0.035^2 \cdot 0.47}{0.58} \approx 0.00187,$$

donde 1.2 kg/m^3 es la densidad del aire, $\pi 0.035^2 \text{ m}^2$ es la superficie transversal de la pelota, 0.47 es una constante de rozamiento que depende de la forma de la pelota (en este caso es la constante asociada a un objeto esférico) y 0.58 kg es la masa de la pelota. Para resolver la ecuación separamos las variables:

$$\frac{dv}{g + kv^2} = -dt.$$

El miembro izquierdo lo vamos a integrar aplicando la regla del arco tangente, para lo cual tenemos que ajustar las constantes:

$$\begin{aligned} \frac{1}{g} \int \frac{dv}{1 + (\sqrt{k/g}v)^2} &= - \int dt, \\ \frac{1}{\sqrt{kg}} \int \frac{\sqrt{k/g} dv}{1 + (\sqrt{k/g}v)^2} &= - \int dt, \\ \frac{1}{\sqrt{kg}} \arctan\left(\sqrt{\frac{k}{g}} v\right) &= -t + c, \\ \arctan\left(\sqrt{\frac{k}{g}} v\right) &= -\sqrt{kg}t + c. \end{aligned}$$

Notemos que en el último paso tendríamos que haber puesto $+\sqrt{kg}c$, pero, como c es una constante arbitraria, podemos pasar a llamar c a $\sqrt{kg}c$. La constante c hay que elegirla de modo que $v(0) = 10 \text{ m/s}^2$:

$$c = \arctan\left(\sqrt{\frac{k}{g}} 10\right) \approx 0.137268.$$

Así, el instante t en el que la velocidad se anula (el momento en la que la pelota llega a su altura máxima) es

$$t^* = \frac{c}{\sqrt{kg}} \approx 1.014 \text{ s}.$$

Por otro lado, despejando v queda:

$$\frac{dh}{dt} = v = \sqrt{\frac{g}{k}} \tan(-\sqrt{kg}t + c) = \frac{1}{k} \frac{\sqrt{kg} \operatorname{sen}(-\sqrt{kg}t + c)}{\cos(-\sqrt{kg}t + c)}.$$

Por lo tanto, la altura de la pelota en función del tiempo es

$$h(t) = \int \frac{1}{k} \frac{\sqrt{kg} \operatorname{sen}(-\sqrt{kg}t + c)}{\cos(-\sqrt{kg}t + c)} = \frac{1}{k} \log \cos(-\sqrt{kg}t + c) + C,$$

donde la constante C hay que elegirla de modo que $h(0) = 0$:

$$C = -\frac{1}{k} \log \cos c \approx 5.054$$

y así concluimos que la altura máxima a la que llega la pelota es

$$h(t^*) = \frac{1}{k} \log \cos(-\sqrt{kg}t^* + c) + C \approx 5.054 \text{ m.}$$

Recordando que, sin rozamiento, la altura máxima era de 5.1 m, vemos que el rozamiento ha reducido esta altura máxima en unos 4.6 cm.

A partir del momento en que la pelota empieza a caer, la ecuación diferencial cambia, pues ahora la velocidad es negativa y el rozamiento positivo:

$$\frac{dv}{dt} = -g + kv^2 \text{ m/s}^2.$$

El cambio de signo hace que ya no se pueda aplicar la regla del arco tangente, pero en su lugar se puede usar la del argumento de la tangente hiperbólica:

$$\begin{aligned} \frac{dv}{g - kv^2} &= -dt, \\ \frac{1}{g} \int \frac{dv}{1 - (\sqrt{k/g}v)^2} &= - \int dt, \\ \frac{1}{\sqrt{kg}} \int \frac{\sqrt{k/g} dv}{1 - (\sqrt{k/g}v)^2} &= - \int dt, \\ \frac{1}{\sqrt{kg}} \operatorname{arg} \tanh\left(\sqrt{\frac{k}{g}} v\right) &= -t + c, \\ \operatorname{arg} \tanh\left(\sqrt{\frac{k}{g}} v\right) &= -\sqrt{kg}t + c. \end{aligned}$$

Si, por simplicidad, empezamos a contar el tiempo de nuevo desde que la pelota empieza a descender, entonces $c = 0$ y

$$v = -\sqrt{\frac{g}{k}} \tanh(\sqrt{kg}t) = -\frac{1}{k} \frac{\sqrt{kg} \operatorname{senh}(\sqrt{kg}t)}{\cosh(\sqrt{kg}t)}.$$

Por lo tanto,

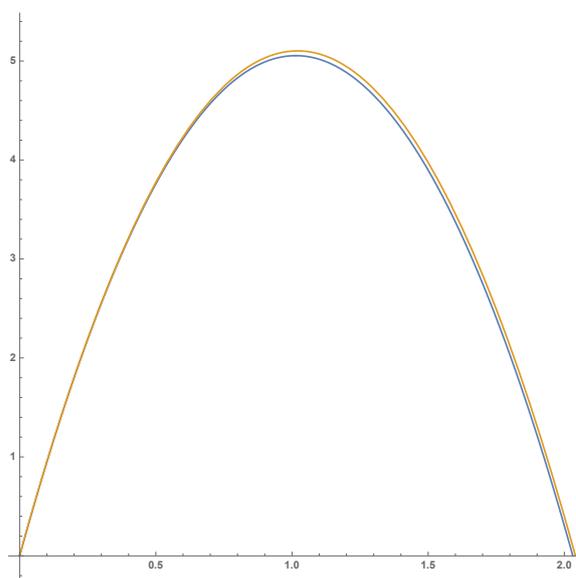
$$h(t) = -\frac{1}{k} \int \frac{\sqrt{kg} \operatorname{senh}(\sqrt{kg}t)}{\cosh(\sqrt{kg}t)} = -\frac{1}{k} \log \cosh(\sqrt{kg}t) + c,$$

donde la constante c tiene que elegirse de modo que $h(0) = 5.054$ m, para lo cual simplemente $c = 5.054$. Así podemos calcular el instante en el que la pelota vuelve al punto de partida:

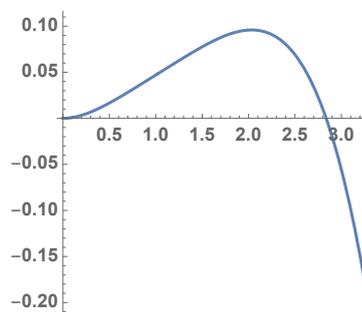
$$\frac{1}{k} \log \cosh(\sqrt{kg} t) = 5.054,$$

que tiene solución $t = 1.01719$, pero a este tiempo le tenemos que sumar el tiempo que la pelota ha tardado en subir, lo que nos da un tiempo total de $t = 2.03$ s, una centésima de segundo menos que sin rozamiento.

La figura siguiente muestra las funciones $h(t)$ correspondientes que resultan de tener en cuenta el rozamiento del aire y de no tenerlo en cuenta:

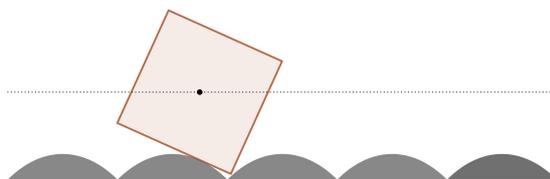


Vemos que el ascenso es muy igualado, pues el rozamiento sólo empieza a notarse cuando la velocidad es suficientemente grande. La gráfica siguiente muestra la diferencia de altura entre la pelota sin rozamiento y la pelota con rozamiento:

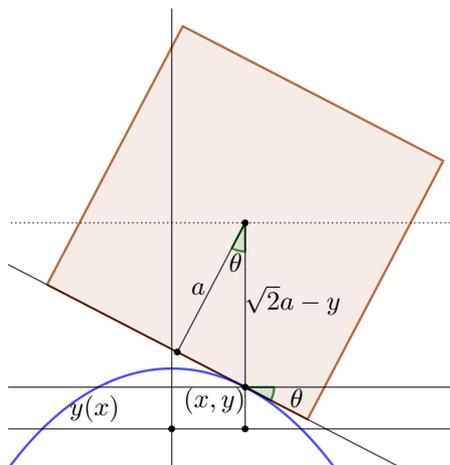


En el momento en que la pelota sin rozamiento llega al punto de partida, la pelota con rozamiento está unos 9.6 cm más alta, pero si siguen cayendo, al poco tiempo la pelota sin rozamiento empieza a reducir la distancia y finalmente sobrepasa a la otra. ■

Ejemplo: Ruedas cuadradas Una bicicleta con ruedas cuadradas no es muy práctica. La razón es que si, por ejemplo, la rueda tiene sus lados de longitud 2, cuando uno de los lados descansa sobre el suelo, el eje de la rueda está a altura 1, pero cuando la rueda tiene una diagonal paralela al suelo, la altura del eje tiene que ser $\sqrt{2}$. Esto significa que el ciclista está continuamente subiendo y bajando, pero además esto significa que, al pedalear, no sólo tiene que hacer la fuerza necesaria para que la bicicleta avance, sino también la necesaria para elevar su peso una altura $\sqrt{2} - 1$, para al cabo de un momento caer bruscamente esa misma distancia y tener que volver a subirla. Sin embargo, vamos a ver ahora que sobre un suelo cuidadosamente diseñado una bicicleta con ruedas cuadradas puede avanzar suavemente, con el eje siempre a la misma altura, y sin más esfuerzo que el que requiere una bicicleta ordinaria:



Concretamente, si la rueda tiene lado $2a$, queremos diseñar el suelo con “montañitas” de modo que la rueda se apoye sobre una esquina siempre que pase de montaña a montaña. En ese momento la altura del eje respecto a la base de las montañas será el radio de la rueda, es decir, $\sqrt{2}a$, y queremos que el eje se mantenga siempre a esa misma altura. Cada “montaña” será la gráfica de una función $y(x)$ y queremos que el punto de tangencia (x, y) de la rueda y la gráfica esté sobre la vertical del eje:



Se forma entonces un triángulo rectángulo de hipotenusa $\sqrt{2}a - y$ y un cateto de longitud a , cuyo ángulo θ en el eje de la rueda es el mismo que forma el lado con la horizontal. Como dicho lado tiene que ser tangente a la gráfica de $y(x)$, lo que queremos es que $\tan \theta = y'(x)$. Por consiguiente,

$$\frac{1}{1 + y'(x)^2} = \cos^2 \theta = \frac{a^2}{(\sqrt{2}a - y)^2}.$$

Despejando:

$$\frac{dy}{dx} = \pm \sqrt{(\sqrt{2} - y/a)^2 - 1}.$$

Esto es una ecuación diferencial que podemos resolver separando las variables:

$$\pm \frac{dy}{\sqrt{(\sqrt{2} - y/a)^2 - 1}} = dx.$$

Integramos y ajustamos una constante para aplicar la regla del argumento del coseno hiperbólico:

$$\mp a \int \frac{(-1/a)dy}{\sqrt{(\sqrt{2} - y/a)^2 - 1}} = \int dx,$$

con lo que

$$\mp a \operatorname{arcosh}(\sqrt{2} - y/a) = x + c,$$

luego

$$\operatorname{arcosh}(\sqrt{2} - y/a) = \mp x/a + c.$$

Podemos exigir que la cumbre de la montaña esté en $x = 0$, lo que se traduce en que $y(0) = (\sqrt{2} - 1)a$, y a su vez en que $c = \operatorname{arcosh}(1) = 0$. Así pues:

$$\sqrt{2} - y/a = \cosh(x/a),$$

donde hemos quitado el signo \mp porque $\cosh(-x) = \cosh x$. Concluimos que

$$y = \sqrt{2}a - a \cosh(x/a).$$

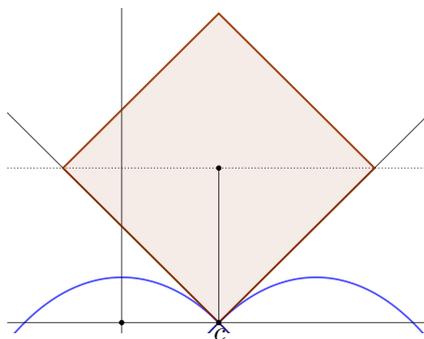
Sólo nos interesa la parte de esta función que queda sobre el eje de abscisas, luego los extremos de la “montaña” son las soluciones de $\cosh(x/a) = \sqrt{2}$, es decir, son los puntos $\pm c$, con

$$c = a \operatorname{arcosh} \sqrt{2}.$$

Notemos que la derivada de $y(x)$ en estos puntos es $-\operatorname{senh}(\pm \operatorname{arcosh}(\sqrt{2}))$ y, teniendo en cuenta la relación fundamental $\cosh^2 x - \operatorname{senh}^2 x = 1$, vemos que cuando $\cosh x = \sqrt{2}$ se cumple que $\operatorname{senh}^2 x = 1$, luego

$$y'(\pm c) = \mp 1.$$

Esto significa que si encadenamos traslaciones de la función $y(x)$, en el punto de corte de dos “montañas” la que acaba tiene derivada -1 y la que empieza tiene derivada 1 , con lo que el cuadrado encaja perfectamente:



Observemos además que la longitud de cada “montaña” es

$$\begin{aligned} L &= \int_{-c}^c \sqrt{1 + \sinh^2(x/a)} dx = \int_{-c}^c \cosh(x/a) dx = a \int_{-c}^c \frac{1}{a} \cosh(x/a) dx \\ &= a[\sinh(x/a)]_{-c}^c = 2a \sinh(c/a) = 2a, \end{aligned}$$

pues si $\cosh(c/a) = \sqrt{2}$, la relación $\cosh^2(c/a) - \sinh^2(c/a) = 1$ nos da que $\sinh(c/a) = 1$.

Así pues, la longitud del arco es igual al lado del cuadrado, como se deduce también del hecho de que, con el movimiento de la rueda, dicho lado va “midiendo” el arco de extremo a extremo. ■

Una caracterización de la lemniscata de Bernoulli Veamos ahora que la ecuación diferencial

$$\rho' = \rho \tan 2\theta$$

caracteriza las coordenadas polares de las lemniscatas de Bernoulli. En efecto, separando las variables e integrando queda

$$\int \frac{d\rho}{\rho} = \int \frac{\sen 2\theta}{\cos 2\theta} d\theta,$$

de donde

$$\log \rho = \frac{1}{2} \log \cos 2\theta + c,$$

luego, llamando $a = e^c$,

$$\rho = ae^{(1/2) \log \cos 2\theta} = a\sqrt{\cos 2\theta},$$

que es, en efecto, la ecuación en coordenadas polares de la lemniscata de Bernoulli presentada en la sección B.5.

Similarmente, la ecuación

$$\rho' = \frac{\rho}{\tan 2\theta} \tag{2.10}$$

tiene por soluciones las funciones $\rho = a\sqrt{\sen 2\theta}$, que corresponden a lemniscatas giradas -45° , pues en coordenadas polares un giro consiste simplemente en cambiar θ por $\theta - \pi/4$, con lo que en el caso de la lemniscata obtenemos

$$\rho = a\sqrt{\cos(2(\theta - \pi/4))} = a\sqrt{\cos(\pi/2 - 2\theta)} = a\sqrt{\sen 2\theta}. \quad \blacksquare$$

Una caracterización de la cicloide Vamos a resolver la ecuación diferencial

$$y(1 + y'^2) = 2R,$$

donde $R > 0$ es una constante. Toda solución tiene que cumplir $y(x) > 0$, de donde a su vez cumple $0 < y(x) \leq 2R$. Despejamos la derivada:

$$\frac{dy}{dx} = \pm \sqrt{\frac{2R - y}{y}} \quad (2.11)$$

y separamos las variables:

$$dx = \pm \sqrt{\frac{y}{2R - y}} dy,$$

con lo cual suponemos que $y(x) < 2R$.

La integral de la derecha la tenemos resuelta en (503) mediante el cambio de variable $y(t') = 2R \sin^2 t'$, con $0 < t' < \pi/2$, pero ahora nos conviene hacer $t' = t/2$, con lo que el cambio es entonces

$$y(t) = R(1 - \cos t), \quad 0 < t < \pi.$$

Si escribimos el resultado de la integral sin deshacer el cambio de variable queda

$$x = \pm(2Rt' - 2R \sin t' \cos t') + c = \pm R(t - \sin t) + c,$$

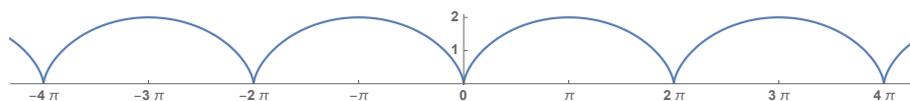
para $0 < t < \pi$. Con esto hemos obtenido que las funciones $y(x)$ que cumplen la ecuación diferencial pueden expresarse en la forma

$$(x, y(x)) = (\pm R(t - \sin t) + c, R(1 - \cos t)), \quad 0 < t < \pi.$$

Si hacemos $c = 0$ y tomamos el signo positivo, obtenemos

$$r(t) = R(t - \sin t, 1 - \cos t),$$

que no es sino la cicloide descrita en la sección B.1:



Lo que hemos probado es que las soluciones de la ecuación con signo positivo son las traslaciones horizontales de la cicloide $r(t)$ restringida al intervalo $]0, \pi[$ (donde $y'(x) > 0$), mientras que las soluciones con signo negativo son las traslaciones de la curva

$$R(-t + \sin t, 1 - \cos t) = r(-t),$$

también con $0 < t < \pi$ o, lo que es lo mismo, las traslaciones de $r(t)$ con $-\pi < t < 0$. Ahora bien, en la gráfica se ve (y es fácil comprobarlo), que la

restricción de r al intervalo $\pi < t < 2\pi$ es una traslación de la restricción a $-\pi < t < 0$, por lo que podemos afirmar que todas las soluciones de la ecuación diferencial son de la forma

$$(x, y(x)) = (R(t - \operatorname{sen} t) + c, R(1 - \operatorname{cos} t)), \quad 0 < t < 2\pi.$$

Con esto hemos recuperado el punto $(R\pi, 2R)$ correspondiente a $t = \pi$, en el que se cumple $y = 2R$, $y' = 0$, luego también cumple la ecuación diferencial (aunque lo habíamos perdido al despejar).

En suma, las soluciones de la ecuación diferencial son las funciones $y(x)$ cuya gráfica resulta de trasladar el arco de cicloide correspondiente a $0 < t < 2\pi$, lo cual incluye en particular a cada uno de los demás arcos¹¹ de $r(t)$.

Si consideramos concretamente la función $y(x)$ correspondiente a $r(t)$, es decir, la que cumple $y(0) = 0$, no podemos decir que cumpla la ecuación en los puntos $2k\pi$ porque en ellos no es derivable. En efecto, en la gráfica se ve que la tangente es vertical, y también se puede deducir a partir de la igualdad (2.11), que muestra que cuando $x \rightarrow 0$ (y, por consiguiente, $y \rightarrow 0$) la derivada tiende a $\pm\infty$.

En conclusión, podemos decir que la función $y : [0, 2R\pi] \rightarrow \mathbb{R}$ cuya gráfica es $r(t)$ para $0 \leq t \leq 2\pi$ es la única solución de la ecuación diferencial dada que cumple las *condiciones de frontera* $y(0) = y(2R\pi) = 0$, en el sentido de que es continua en $[0, 2R\pi]$, es derivable en $]0, 2R\pi[$, en el intervalo abierto es solución de la ecuación diferencial y además cumple las condiciones de frontera. (En realidad basta imponer $y(0) = 0$.) Cualquier otra solución es una traslación horizontal de ésta.

En otras palabras, la solución de la ecuación diferencial con condición de frontera

$$y(1 + y'^2) = 2R, \quad y(0) = 0$$

es un arco de cicloide de radio R . ■

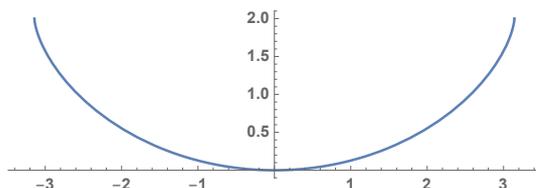
Otra caracterización de la cicloide Consideremos la cicloide

$$r(t) = R(t + \operatorname{sen} t, 1 + \operatorname{cos} t),$$

parametrizada de modo que tiene su máximo en $t = 0$. Podemos ponerla “boca arriba” aplicándole la simetría $y \mapsto 2R - y$ respecto a la recta horizontal $y = R$, con lo que la parametrización pasa a ser:

$$r(t) = R(t + \operatorname{sen} t, 1 - \operatorname{cos} t).$$

¹¹En toda esta discusión es crucial que exigimos que el dominio de una solución de una ecuación diferencial sea un intervalo abierto, y no una unión de intervalos abiertos disjuntos, pues en tal caso las restricciones de la solución a cada intervalo no guardan ninguna relación entre sí, por lo que no es razonable considerarlas parte de “la misma solución”.



En estos términos $r'(t) = R(1 + \cos t, \sin t)$ y

$$\|r'(t)\| = R\sqrt{2(1 + \cos t)} = 2R \cos(t/2).$$

Por lo tanto, la función $s(t)$ que da la longitud del arco de cicloide es

$$s(t) = 2R \int_0^t \cos \frac{t}{2} dt = 4R \sin \frac{t}{2} = 4R \sqrt{\frac{1 - \cos t}{2}}.$$

Comparando con la coordenada $y(t) = R(1 - \sin t)$ de la parametrización, vemos que

$$y = \frac{1}{8R} s^2.$$

Así pues:

La altura de un punto una cicloide “invertida” es proporcional al cuadrado de su distancia al punto mínimo medida sobre la curva.

Vamos a demostrar que esta propiedad caracteriza a la cicloide, es decir:

Si una curva tiene altura mínima en un punto y la altura de cada punto sobre dicha altura mínima es proporcional al cuadrado de la distancia sobre la curva al punto mínimo, necesariamente se trata de una cicloide.

Para probarlo consideramos una curva arbitraria $y(x)$ (o $r(x) = (x, y(x))$), con un mínimo en $y(0) = 0$, y de modo que $y = s^2/8R$, para cierta constante $R > 0$, donde $s(x)$ es la función que da la longitud de arco desde $x = 0$. Así,

$$\frac{ds}{dx} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}.$$

Por otro lado, derivando la relación dada:

$$\frac{dy}{dx} = \frac{s}{4R} \frac{ds}{dx} = \frac{\sqrt{8Ry}}{4R} \frac{ds}{dx} = \sqrt{\frac{y}{2R}} \frac{ds}{dx}$$

Sustituyendo queda

$$\frac{dy}{dx} = \sqrt{\frac{y}{2R}} \sqrt{1 + \left(\frac{dy}{dx}\right)^2},$$

luego

$$\left(\frac{dy}{dx}\right)^2 = \frac{y}{2R} \left(1 + \left(\frac{dy}{dx}\right)^2\right),$$

luego

$$\frac{dy}{dx} = \pm \sqrt{\frac{y/2R}{1 - y/2R}},$$

donde el doble signo se debe a que, como $y(x)$ tiene un mínimo en 0, la derivada tiene que ser negativa para $x < 0$ y positiva para $x > 0$.

Tenemos así una ecuación diferencial de variables separables:

$$\pm \sqrt{\frac{1 - y/2R}{y/2R}} dy = dx,$$

que nos lleva a

$$x = \pm \int_0^{y(x)} \sqrt{\frac{1 - y/2R}{y/2R}} dy,$$

Para calcular la integral hacemos el cambio $u = \sqrt{y/2R}$, de modo que $y = 2Ru^2$ y $dy = 4Ru du$:

$$x = \pm \int_0^{\sqrt{y(x)/2R}} \frac{\sqrt{1 - u^2}}{u} 4Ru du = \pm 4R \int_0^{\sqrt{y(x)/2R}} \sqrt{1 - u^2} du.$$

Esta integral la hemos calculado¹² en (C.7) y en (C.14):

$$x = \pm 4R \left[\frac{u}{2} \sqrt{1 - u^2} + \frac{1}{2} \arcsen u \right]_0^{\sqrt{y(x)/2R}},$$

luego

$$x = \pm 4R \left(\frac{1}{2} \sqrt{\frac{y}{2R}} \sqrt{1 - \frac{y^2}{4R^2}} + \frac{1}{2} \arcsen \sqrt{\frac{y}{2R}} \right).$$

Aunque no lo parezca, ésta es la relación entre las coordenadas de una cicloide. Para ponerlo en evidencia llamamos $t/2 = \arcsen \sqrt{y/2R}$, de modo que

$$y = 2R \sen^2 \frac{t}{2} = R(1 - \cos t).$$

Por otra parte,

$$x = \pm 4R \left(\frac{1}{2} \sen \frac{t}{2} \cos \frac{t}{2} + \frac{t}{4} \right) = \pm 4R \left(\frac{1}{4} \sen t + \frac{t}{4} \right) = \pm R(t + \sen t).$$

¹²Hemos cambiado $-\arccos u$ por $\arcsen u$, pues ambas funciones tienen la misma derivada.

Notemos que, en principio, $t/2$ tiene seno positivo, luego varía entre 0 y π , pero permitiendo que t varíe entre $-\pi$ y π incorporamos el doble signo, por lo que podemos escribir simplemente:

$$(x(t), y(t)) = (R(t + \operatorname{sen} t), R(1 - \operatorname{cost})),$$

que es una de las parametrizaciones naturales de una cicloide de radio R . ■

2.9 El teorema de Arzelà

En esta sección demostraremos un resultado cuya prueba es un tanto técnica y que, en realidad, no vamos a necesitar más que para no introducir hipótesis innecesarias en algunos teoremas posteriores. Se trata de una generalización del teorema 2.8 que aumenta considerablemente la potencia del rudimentario cálculo integral —comparado con la teoría de integración moderna— que hemos expuesto. He aquí su enunciado:

Teorema 2.16 (Teorema de Arzelà) *Sea $\{f_n\}_{n=0}^{\infty}$ una sucesión de funciones continuas $f_n : [a, b] \rightarrow \mathbb{R}$ uniformemente acotadas, es decir, tales que existe un $C > 0$ tal que $|f_n(x)| \leq C$ para todo $x \in [a, b]$. Sea $f : [a, b] \rightarrow \mathbb{R}$ otra función continua tal que, para todo $x \in [a, b]$ se cumpla $\lim_n f_n(x) = f(x)$. Entonces*

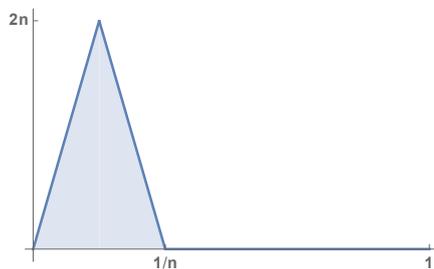
$$\lim_n \int_a^b f_n(x) dx = \int_a^b f(x) dx.$$

Notemos que 2.8 es un caso particular de este teorema, pues si la sucesión es monótona, podemos tomar como C el máximo de una cota de f_0 y una de f . La existencia de C es necesaria, como se pone de manifiesto considerando las funciones f_n cuya gráfica es un triángulo isósceles de base $[1, 1/n]$ y altura $2n$ (véase la figura siguiente), de modo que, para todo $x \in [0, 1]$, se cumple

$$\lim_n f_n(x) = 0,$$

pero

$$\int_0^1 f_n(x) dx = 1 \neq \int_0^1 0 dx.$$



Para probar el teorema necesitamos algunos resultados básicos sobre figuras elementales en \mathbb{R} , análogas a las que definimos en la sección A.2 de [ITan] en \mathbb{R}^2 para definir la medida de Jordan.

Definición 2.17 Una *celda* C en \mathbb{R} es un intervalo de una de las formas

$$]a, b[, \quad [a, b[, \quad]a, b], \quad [a, b],$$

incluyendo los casos $[a, b] = \{a\}$ y $]a, a[= \emptyset$. En cualquiera de los casos, definimos el *contenido* de una celda C como $|C| = b - a$. Una *figura elemental* es una unión finita de celdas.

Simplificando la prueba del teorema [ITAn A.2] se demuestra:

Teorema 2.18 Si A y B son figuras elementales en \mathbb{R} , también lo son

$$A \cup B, \quad A \cap B, \quad A \setminus B.$$

Es claro que si A es una figura elemental en \mathbb{R} , entonces $A \times [0, 1]$ es una figura elemental en \mathbb{R}^2 en el sentido definido en [ITAn A.1], por lo que podemos definir su *medida* como $m(A) = m(A \times [0, 1])$, donde la segunda medida es la definida en [ITAn A.6], de modo que la medida que acabamos de definir es ciertamente una medida finitamente aditiva en el sentido de [ITAn A.5], y cumple todas las propiedades que se prueban a continuación de dicha definición. Observemos que la medida de una celda es su contenido.

Teorema 2.19 Sea $\{A_n\}_{n=1}^{\infty}$ una sucesión decreciente de subconjuntos acotados de \mathbb{R} cuya intersección sea vacía (es decir, que ningún punto pertenezca a todos ellos). Sea

$$s_n = \sup\{m(E) \mid E \subset A_n \text{ es elemental}\}.$$

Entonces $\lim_n s_n = 0$.

DEMOSTRACIÓN: Observemos que $s_{n+1} \leq s_n$. Si la sucesión no converge a 0, sustituyéndola por una subsucesión podemos suponer que existe un $\epsilon > 0$ tal que $s_n > \epsilon$ para todo n . Entonces, para cada n , existe $E_n \subset A_n$ elemental tal que $m(E_n) > s_n - \epsilon/2^n$. Reduciendo si es preciso los intervalos que forman E_n podemos suponer que todos ellos son cerrados.

Sea $H_n = E_1 \cap \dots \cap E_n$ y vamos a demostrar que no es el conjunto vacío. Admitiendo esto de momento, podemos tomar $x_n \in H_n$ y, por el teorema de Bolzano-Weierstrass [ITAn 1.24], la sucesión $\{x_n\}_{n=1}^{\infty}$ tiene una subsucesión convergente, digamos $\lim_n x_{n_k} = x^*$. Fijado un índice n , si $n_k \geq n$, entonces $x_{n_k} \in H_{n_k} \subset E_n$, luego E_n contiene infinitos términos de la subsucesión. Además es una unión finita de intervalos cerrados, luego uno de ellos tiene que contener infinitos términos de la subsucesión, luego el límite x^* también está en dicho intervalo, luego $x^* \in E_n \subset A_n$ para todo n , cuando estamos suponiendo que los A_n tienen intersección vacía.

Así pues, basta probar que H_n no es vacío. Para ello observamos que si $E \subset A_n \setminus E_n$ es elemental, entonces, usando las propiedades de las medidas finitamente aditivas, $m(E) + m(E_n) = m(E \cup E_n) \leq s_n$ y $m(E_n) > s_n - \epsilon/2^n$, luego $m(E) < \epsilon/2^n$.

Por lo tanto, si $E \subset A_n \setminus H_n$ es elemental,

$$E = (E \setminus E_1) \cup (E \setminus E_2) \cup \cdots \cup (E \setminus E_n)$$

y, como $E \setminus E_i \subset A_i \setminus E_i$, hemos probado que $m(E \setminus E_i) < \epsilon/2^i$, luego

$$m(E) \leq \frac{\epsilon}{2} + \frac{\epsilon}{4} + \cdots + \frac{\epsilon}{2^n} < \epsilon.$$

Finalmente, como $s_n > \epsilon$, tiene que haber un $E \subset A_n$ elemental de modo que $m(E) > \epsilon$, y hemos probado que dicho E no está contenido en $A_n \setminus H_n$, luego E tiene puntos en H_n , luego éste no es vacío. ■

DEMOSTRACIÓN (de 2.16): Sustituyendo f_n por $f_n - f$ podemos suponer que $f(x) = 0$, para todo $x \in [a, b]$. Por otra parte, $f_n(x) = f_n^+(x) - f_n^-(x)$, donde f_n^+ y f_n^- son las funciones (continuas) definidas antes del teorema 2.4. Claramente ambas convergen a 0 igual que f , luego basta probar el teorema para cada una de ellas. Alternativamente, podemos suponer que $f_n(x) \geq 0$, para todo $x \in [a, b]$.

Definimos A_n como el conjunto de los puntos de $[a, b]$ tales que existe un $i \geq n$ para el cual $f_i(x) \geq \epsilon/(b-a)$. Claramente $A_{n+1} \subset A_n$ y el hecho de que la sucesión de funciones tiende a 0 implica que la intersección de todos los A_n es vacía. El teorema anterior nos da entonces que, dado $\epsilon > 0$, existe un N tal que si $n \geq N$, entonces $s_n < \epsilon/2C$, con lo que si $E \subset A_n$ es elemental, $m(E) < \epsilon/2C$. Basta probar que si $n \geq N$ se cumple que

$$\int_a^b f_n(x) dx < \epsilon.$$

En caso contrario tomamos δ en las condiciones del teorema 2.5 tomando como ϵ para dicho teorema el número

$$\epsilon - \int_a^b f_n(x) dx > 0.$$

Tomamos una partición

$$a = x_0 < x_1 < \cdots < x_n < b$$

tal que $\Delta x_i = x_i - x_{i-1} < \delta$ y tomamos como ξ_i el punto del intervalo $[x_{i-1}, x_i]$ donde $f_n(\xi_i)$ toma su valor mínimo. Entonces

$$\int_a^b f_n(x) dx - \sum_{i=1}^n f_n(\xi_i) \Delta x_i < \epsilon - \int_a^b f_n(x) dx,$$

luego

$$\epsilon < \sum_{i=1}^n f(\xi_i) \Delta x_i.$$

Sea E la unión de los intervalos $[x_{i-1}, x_i]$ tales que $f(\xi_i) \geq \epsilon/2(b-a)$, con lo que es una figura elemental $E \subset A_n$, luego sabemos que $m(E) < \epsilon/2C$. Descomponemos

$$\epsilon < \sum_{i=1}^n f(\xi_i) \Delta x_i = \sum_i f(\xi_i) \Delta x_i + \sum_j f(\xi_j) \Delta x_j \leq Cm(E) + \frac{\epsilon}{2(b-a)}(b-a) < \epsilon,$$

donde i recorre los índices correspondientes a intervalos $[x_{i-1}, x_i] \subset E$ y j los índices restantes, y así tenemos una contradicción. ■

Capítulo III

Integrales impropias

En el capítulo anterior nos hemos encontrado ya algunos ejemplos de lo que se conoce como “integrales impropias”, concretamente, al calcular el volumen y el área del “cuerno de Gabriel” vimos que dicho volumen es

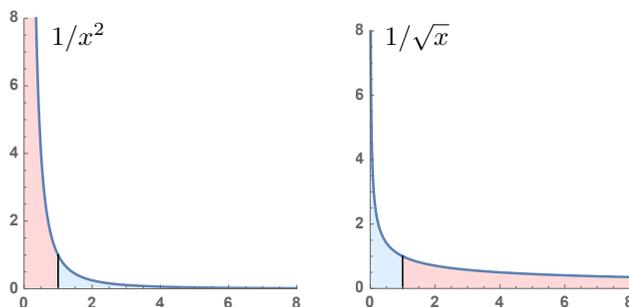
$$V = \int_1^{+\infty} \frac{\pi}{x^2} dx = \lim_{x_0 \rightarrow +\infty} \int_1^{x_0} \frac{\pi}{x^2} dx = \lim_{x_0 \rightarrow +\infty} \pi - \frac{\pi}{x_0} = \pi.$$

Aquí la segunda igualdad se entiende como una definición: integrar una función desde 1 hasta $+\infty$ significa integrar desde 1 hasta un punto arbitrario x_0 y luego calcular el límite cuando x_0 tiende a $+\infty$. En el caso concreto del “cuerno de Gabriel”, la integral hasta x_0 se interpreta como el volumen del cuerno “truncado” a una altura x_0 , de modo que la integral hasta $+\infty$ se interpreta como el volumen del cuerno completo.

En este capítulo vamos a estudiar este tipo de integrales en intervalos infinitos que involucran el cálculo de un límite. Si consideramos la interpretación más directa de una integral como un área en lugar de como un volumen y eliminamos π del ejemplo precedente, el cálculo

$$\int_1^{+\infty} \frac{1}{x^2} dx = \lim_{t \rightarrow +\infty} \int_1^t \frac{1}{x^2} dx = \lim_{t \rightarrow +\infty} \left[-\frac{1}{x} \right]_1^t = \lim_{t \rightarrow +\infty} 1 - \frac{1}{t} = 1$$

se interpreta como que el área marcada en la figura de la izquierda sobre el intervalo $[1, +\infty[$ es de 1 unidad cuadrada.



Podría haber sido infinita, que es lo que sucede, de hecho, si consideramos la función $1/\sqrt{x}$ (representada a la derecha):

$$\int_1^{+\infty} \frac{1}{\sqrt{x}} dx = \lim_{t \rightarrow +\infty} \int_1^t \frac{1}{\sqrt{x}} dx = \lim_{t \rightarrow +\infty} [2\sqrt{x}]_1^t = \lim_{t \rightarrow +\infty} 2\sqrt{t} - 2 = +\infty.$$

Pero hay otro sentido en el que una integral puede ser “impropia”, y es el que surge si nos planteamos si el área que queda bajo la gráfica de $1/x^2$ (o de $1/\sqrt{x}$) sobre el intervalo $[0, 1]$ es finita o infinita. Ninguna de las dos está definida en $x = 0$, pues ambas tienden a $+\infty$ en 0, por lo que no tenemos garantías de que dicha área tenga que ser finita, pero podría serlo.

Por ejemplo, para la función $1/x^2$ podemos calcular

$$\int_t^1 \frac{1}{x^2} dx = \left[-\frac{1}{x} \right]_t^1 = \frac{1}{t} - 1,$$

y esta expresión nos da el área situada sobre el intervalo $[t, 1]$, por lo que el área sobre todo el intervalo $[0, 1]$ será

$$\int_0^1 \frac{1}{x^2} dx = \lim_{t \rightarrow 0^+} \int_t^1 \frac{1}{x^2} dx = \lim_{t \rightarrow 0^+} \frac{1}{t} - 1 = +\infty.$$

En este caso el área es infinita, pero en el caso de $1/\sqrt{x}$ es finita:

$$\int_0^1 \frac{1}{\sqrt{x}} dx = \lim_{t \rightarrow 0^+} \int_t^1 \frac{1}{\sqrt{x}} dx = \lim_{t \rightarrow 0^+} 2 - 2\sqrt{t} = 2.$$

Vamos a dar definiciones precisas de estos dos tipos de integrales “impropias” y de las integrales “mixtas” que son impropias en sus dos extremos.

3.1 Definición y ejemplos

Definición 3.1 Si $f : [a, +\infty[\rightarrow \mathbb{R}$ es una función continua, definimos

$$\int_a^{+\infty} f(x) dx = \lim_{t \rightarrow +\infty} \int_a^t f(x) dx.$$

A esto lo llamamos una *integral impropia de primera especie*, que será *convergente* si el límite existe y es finito, y *divergente* en caso contrario. Igualmente, para una función continua $f :]-\infty, b] \rightarrow \mathbb{R}$ definimos

$$\int_{-\infty}^b f(x) dx = \lim_{t \rightarrow -\infty} \int_t^b f(x) dx.$$

Por otra parte, si $f : [a, b[\rightarrow \mathbb{R}$ es una función continua, definimos

$$\int_a^b f(x) dx = \lim_{t \rightarrow b^-} \int_a^t f(x) dx,$$

y diremos que se trata de una *integral impropia* (en b) *de segunda especie*. Diremos que es *convergente* si el límite existe y es finito, y en caso contrario diremos que es *divergente*. Similarmente, si $f :]a, b[\rightarrow \mathbb{R}$ es continua, definimos

$$\int_a^b f(x) dx = \lim_{t \rightarrow a^+} \int_t^b f(x) dx.$$

Por último, si $f :]a, b[\rightarrow \mathbb{R}$ es una función continua, donde a puede ser $-\infty$ y b puede ser $+\infty$, definimos

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx,$$

entendiendo que la integral es convergente si y sólo si lo son las dos integrales del miembro derecho, donde c es cualquier número real $a < c < b$ (y es fácil ver que la convergencia o divergencia, así como el valor de la integral si es convergente, no dependen de la elección de c).

Observemos que en 2.2 habíamos definido la integral cuando a y b son finitos y el integrando f tiene primitiva en $[a, b]$, mientras que ahora tenemos definida la integral para una función continua arbitraria (y con a y b posiblemente infinitos) en términos de un límite (o una suma de dos límites) que pueden ser infinitos o no existir.

De las propiedades de los límites se sigue inmediatamente que esta integral generalizada sigue cumpliendo las propiedades básicas siguientes:

$$\begin{aligned} \int_a^b (f(x) + g(x)) dx &= \int_a^b f(x) dx + \int_a^b g(x) dx, & \int_a^b cf(x) dx &= c \int_a^b f(x) dx, \\ \int_a^b f(x) dx &= - \int_b^a f(x) dx, & \int_a^b f(x) dx &= \int_a^c f(x) dx + \int_c^b f(x) dx, \end{aligned}$$

entendiendo que los miembros izquierdos son convergentes si lo son las integrales de los miembros derechos, y en tal caso se dan las igualdades.

Ejemplos Conviene observar que las integrales impropias de segunda especie convergentes están contempladas ya en la definición 2.2, donde sólo exigimos que el integrando esté definido en $]a, b[$. Vamos a discutir la situación con algunos ejemplos concretos:

$$\int_0^1 \frac{1}{\sqrt{x}} dx, \quad \int_0^1 x^2 dx, \quad \int_0^1 \frac{1}{\sqrt{x}} e^{-\sqrt{x}} dx, \quad \int_0^1 \log x dx \quad (3.1)$$

Podemos decir que la primera integral es impropia en 0, pues el integrando es una función continua en $]0, 1]$. Esto justifica el cálculo de esta integral que hemos hecho en la introducción de este capítulo. Pero, alternativamente, sin más que aplicar 2.2, podemos argumentar que el integrando tiene primitiva $2\sqrt{x}$ en $[0, 1]$, por lo que el cálculo puede reducirse a

$$\int_0^1 \frac{1}{\sqrt{x}} dx = [2\sqrt{x}]_0^1 = 2,$$

sin necesidad de tomar límites.

El resultado es el mismo y no es casualidad. En general, si tenemos una función continua $f : [a, b[\rightarrow \mathbb{R}$, una primitiva en $[a, b[$ es

$$F(t) = \int_a^t f(x) dx,$$

pero para que la integral esté definida en el sentido de 2.2 es necesario y suficiente que F admita una extensión continua al intervalo $[a, b]$, lo cual a su vez equivale a que exista y sea finito el límite

$$\lim_{t \rightarrow b^-} F(t) = \lim_{t \rightarrow b^-} \int_a^t f(x) dx,$$

en cuyo caso podemos extender F a b mediante

$$F(b) = F(b) - F(a) = \lim_{t \rightarrow b^-} \int_a^t f(x) dx = \int_a^b f(x) dx,$$

donde la última integral es que hemos definido ahora y $F(b) - F(a)$ es el valor de la integral definido en 2.2.

Lo mismo vale para integrales impropias en su extremo inferior y, descomponiendo en suma de dos las integrales impropias de segunda especie en ambos extremos, es fácil concluir que las integrales impropias de segunda especie (en uno o en sus dos extremos) convergentes son exactamente las integrales que están definidas en 2.2. Lo único que aporta la definición de integral impropia de segunda especie es que una integral como

$$\int_0^1 \frac{1}{x^2} dx$$

no está definida según 2.2 y está definida como $+\infty$ como integral impropia.

Por otro lado, consideremos ahora la segunda integral de (3.1). Lo natural es considerar que no es una integral impropia, y calcularla según 2.2 como

$$\int_0^1 x^2 dx = \left[\frac{x^3}{3} \right] = \frac{1}{3}.$$

Sin embargo, técnicamente, nada nos impide considerar al integrando como una función continua en $[0, 1[$ y calcular

$$\int_0^1 x^2 dx = \lim_{t \rightarrow 1^-} \int_0^t x^2 dx = \lim_{t \rightarrow 1^-} \frac{t^3}{3} = \frac{1}{3}.$$

También podríamos haber considerado que es impropia en 0 o en ambos extremos a la vez. Desde un punto de vista conceptual, lo natural sería añadir a la definición de integral impropia de segunda especie en un extremo a o b de su intervalo que el integrando no admita extensión continua a dicho extremo, de tal modo que la integral anterior no sería impropia, pero desde un punto de vista

técnico no hay necesidad de introducir tal exigencia, y es más práctico considerar que las integrales “propias” (es decir, las integrales de funciones continuas en intervalos $[a, b]$) son un caso particular de las impropias de segunda especie convergentes. La utilidad de este convenio se aprecia en el tercer ejemplo en (3.1). En principio el integrando no está definido en 0, por lo que “podría parecer” que la integral es impropia en 0, si bien una consideración más atenta muestra que el integrando se extiende a una función continua en $[0, 1]$ definiendo $f(0) = 0$, pero no necesitamos observar, ni mucho menos justificar que esto es así para calcular la integral como impropia:

$$\begin{aligned} \int_0^1 \frac{1}{\sqrt{x}} e^{-\sqrt{x}} dx &= \lim_{t \rightarrow 0^+} \int_t^1 \frac{1}{\sqrt{x}} e^{-\sqrt{x}} dx = \lim_{t \rightarrow 0^+} [-2e^{-\sqrt{x}}]_t^1 \\ &= \lim_{t \rightarrow 0^+} 2e^{-\sqrt{t}} - 2e^{-1} = 2 - 2/e. \end{aligned}$$

Saber que el integrando se extiende a una función continua en $[0, 1]$ nos garantiza que si vemos la integral como impropia, será convergente. Más aún, podríamos haber aplicado directamente 2.2 y haber prescindido del límite:

$$\int_0^1 \frac{1}{\sqrt{x}} e^{-\sqrt{x}} dx = [-2e^{-\sqrt{x}}]_0^1 = 2 - 2/e.$$

Sin embargo, el cuarto ejemplo muestra que a veces no es inmediato que la primitiva del integrando admita una extensión continua al intervalo cerrado, en cuyo caso la formulación explícita en términos de límites es la más natural. (La primitiva está calculada en (C.4)):

$$\int_0^1 \log x dx = \lim_{t \rightarrow 0^+} \int_t^1 \log x dx = \lim_{t \rightarrow 0^+} [x \log x - x]_t^1 = \lim_{t \rightarrow 0^+} t - t \log t - 1 = -1,$$

donde hemos usado que

$$\lim_{t \rightarrow 0^+} t \log t = \lim_{t \rightarrow 0^+} \frac{\log t}{1/t} = \lim_{t \rightarrow 0^+} \frac{1/t}{-1/t^2} = \lim_{t \rightarrow 0^+} -t = 0. \quad \blacksquare$$

Ejemplo Veamos ahora un ejemplo de integral impropia en sus dos extremos:

$$\int_{-\infty}^{+\infty} \frac{dx}{1+x^2} = \pi.$$

En efecto, para calcular esta integral tenemos que partirla:

$$\int_{-\infty}^{+\infty} \frac{dx}{1+x^2} = \int_{-\infty}^0 \frac{dx}{1+x^2} + \int_0^{+\infty} \frac{dx}{1+x^2},$$

donde hemos elegido arbitrariamente el 0 para la partición. Entonces:

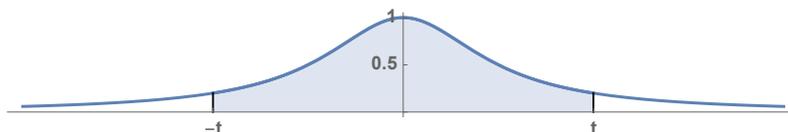
$$\int_{-\infty}^0 \frac{dx}{1+x^2} = \lim_{t \rightarrow -\infty} \int_t^0 \frac{dx}{1+x^2} = \lim_{t \rightarrow -\infty} [\arctan x]_t^0 = \lim_{t \rightarrow -\infty} -\arctan t = \frac{\pi}{2}.$$

Igualmente se concluye que la segunda integral vale $\pi/2$.

Consideremos ahora esta forma alternativa de llegar al mismo resultado:

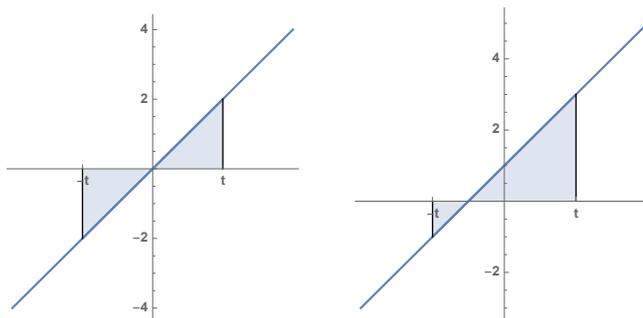
$$\int_{-\infty}^{+\infty} \frac{dx}{1+x^2} = \lim_{t \rightarrow +\infty} \int_{-t}^t \frac{dx}{1+x^2} = \lim_{t \rightarrow +\infty} [\arctan x]_{-t}^t =$$

$$\lim_{t \rightarrow +\infty} \arctan t - \arctan(-t) = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi.$$



La coincidencia no es casual. Al integrar en $[-t, t]$ obtenemos el área sombreada en la figura, y al hacer que t tienda a $+\infty$ obtenemos toda el área situada bajo la gráfica del integrando, igual que si primero integramos en $]-\infty, 0]$ y luego en $[0, +\infty[$. Ahora bien, esta forma de calcular una integral doblemente impropia tiene un inconveniente, que se pone de manifiesto si intentamos aplicarla al ejemplo siguiente:

$$\int_{-\infty}^{+\infty} x \, dx = \lim_{t \rightarrow +\infty} \int_{-t}^t x \, dx = \lim_{t \rightarrow +\infty} \left[\frac{x^2}{2} \right]_{-t}^t = \lim_{t \rightarrow +\infty} 0 = 0.$$



Puede parecer que el resultado es razonable, puesto que con ello estamos permitiendo que, al igual que sucede con las integrales “propias”, las áreas por debajo del eje de abscisas compensan a las que se encuentran por encima y, como muestra la figura de la izquierda, en este caso la función deja un área infinita por debajo del intervalo $]-\infty, 0]$ y otra igual por encima de $[0, +\infty[$, luego es razonable concluir que la integral completa es convergente y vale 0. Sin embargo, no es esto lo que obtenemos si aplicamos la definición de integral impropia, que requiere descomponer la integral como

$$\int_{-\infty}^{+\infty} x \, dx = \int_{-\infty}^0 x \, dx + \int_0^{+\infty} x \, dx,$$

y es claro que ambas integrales son divergentes, luego la integral completa también lo es.

Y sucede que haber ajustado las definiciones para que esta integral sea divergente es más “razonable” que si las definiciones hubieran llevado a que la integral valiera 0. Por ejemplo, si aplicamos el mismo procedimiento de cálculo a la función $x + 1$, cuya gráfica vemos en la figura anterior a la derecha, obtenemos

$$\int_{-\infty}^{+\infty} (x + 1) dx = \lim_{t \rightarrow +\infty} \int_{-t}^t (x + 1) dx = \lim_{t \rightarrow +\infty} \left[\frac{x^2}{2} + x \right]_{-t}^t = \lim_{t \rightarrow +\infty} 2t = +\infty,$$

luego ahora, una función que tiene una parte sobre el eje de abscisas y otra por debajo idénticas a las del caso anterior (meramente trasladadas una unidad hacia la izquierda) no integra 0, sino $+\infty$. Con la definición “correcta”, ambas integrales son divergentes.¹

En general, si $f : \mathbb{R} \rightarrow \mathbb{R}$ es una función continua, se define el *valor principal* de su integral como

$$\text{VP} \int_{-\infty}^{+\infty} f(x) dx = \lim_{t \rightarrow +\infty} \int_{-t}^t f(x) dx.$$

No es difícil probar que si una integral impropia en $]-\infty, +\infty[$ es convergente, entonces coincide con su valor principal, pero si es divergente, el valor principal puede ser finito, como en el ejemplo anterior.

Ahora calcularemos la integral

$$\int_0^{\pi} \log \operatorname{sen} x dx.$$

El primero que calculó esta integral fue Euler en 1769, y durante mucho tiempo se consideró una integral difícil, que para ser calculada con rigor requería emplear funciones de variable compleja. Sin embargo, vamos a ver que admite un cálculo elemental.

Ante todo, como es impropia en ambos extremos, tenemos que partirla:

$$\int_0^{\pi} \log \operatorname{sen} x dx = \int_0^{\pi/2} \log \operatorname{sen} x dx + \int_{\pi/2}^{\pi} \log \operatorname{sen} x dx = 2 \int_0^{\pi/2} \log \operatorname{sen} x dx,$$

porque la segunda integral se convierte en la primera con el cambio $x = \pi - u$. Por otra parte, integrando por partes, vemos que

$$\int_{\epsilon}^{\pi/2} \log \operatorname{sen} x dx = -\epsilon \log \operatorname{sen} \epsilon - \int_{\epsilon}^{\pi/2} \frac{x}{\operatorname{sen} x} \cos x dx,$$

¹En una serie infinita también pueden compensarse términos positivos y negativos, pero eso no significa que tenga sentido “sumar”

$$\cdots (-3) + (-2) + (-1) + 0 + 1 + 2 + 3 + \cdots,$$

que sería algo análogo a calcular la integral $\int_{-\infty}^{+\infty} x dx$.

y el último integrando es una función continua en 0, y además

$$\begin{aligned}\lim_{\epsilon \rightarrow 0} -\epsilon \log \operatorname{sen} \epsilon &= -\lim_{\epsilon \rightarrow 0} \frac{\log \operatorname{sen} \epsilon}{1/\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{\cos \epsilon / \operatorname{sen} \epsilon}{1/\epsilon^2} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\epsilon}{\operatorname{sen} \epsilon} \epsilon \cos \epsilon = 0.\end{aligned}$$

Por lo tanto, la integral es convergente. Para calcularla observamos que

$$I = \int_0^{\pi/2} \log \operatorname{sen} x \, dx = \int_0^{\pi/2} \log \cos x \, dx,$$

porque una integral se transforma en la otra con el cambio $x = \pi/2 - u$. Por lo tanto,

$$\begin{aligned}2I &= \int_0^{\pi/2} (\log \operatorname{sen} x + \log \cos x) \, dx = \int_0^{\pi/2} \log\left(\frac{1}{2} \operatorname{sen} 2x\right) \, dx = \\ &= \int_0^{\pi/2} \log(1/2) \, dx + \int_0^{\pi/2} \log \operatorname{sen} 2x \, dx = -\frac{\pi}{2} \log 2 + \frac{1}{2} \int_0^{\pi} \log \operatorname{sen} t \, dt = \\ &= -\frac{\pi}{2} \log 2 + I,\end{aligned}$$

luego, despejando,

$$\int_0^{\pi/2} \log \operatorname{sen} x \, dx = -\frac{\pi}{2} \log 2,$$

y a su vez,

$$\int_0^{\pi} \log \operatorname{sen} x \, dx = -\pi \log 2. \quad \blacksquare$$

Las integrales de $1/x^2$ y $1/\sqrt{x}$ que hemos calculado son casos particulares de los teoremas siguientes:

Teorema 3.2 Si $0 \leq s < 1$ se cumple que

$$\int_0^1 \frac{1}{x^s} \, dx = \frac{1}{1-s},$$

mientras que si $s \geq 1$ la integral es divergente.

DEMOSTRACIÓN: Si $s \neq 1$ tenemos que

$$\begin{aligned}\int_0^1 \frac{1}{x^s} \, dx &= \lim_{t \rightarrow 0} \int_t^1 \frac{1}{x^s} \, dx = \lim_{t \rightarrow 0} \left[\frac{x^{-s+1}}{-s+1} \right]_t^1 \\ &= \lim_{t \rightarrow 0} \frac{1}{1-s} - \frac{t^{1-s}}{1-s} = \frac{1}{1-s}\end{aligned}$$

si $0 \leq s < 1$, mientras que si $s > 1$ la integral es divergente. Para $s = 1$ la integral se calcula con la regla del logaritmo y también sale divergente. \blacksquare

Similarmente:

Teorema 3.3 Si $s > 1$ se cumple que

$$\int_1^{+\infty} \frac{1}{x^s} dx = \frac{1}{s-1},$$

mientras que si $0 \leq s \leq 1$ la integral es divergente.

DEMOSTRACIÓN: Si $s \neq 1$ tenemos que

$$\begin{aligned} \int_1^{+\infty} \frac{1}{x^s} dx &= \lim_{t \rightarrow +\infty} \int_1^t \frac{1}{x^s} dx = \lim_{t \rightarrow +\infty} \left[\frac{x^{-s+1}}{-s+1} \right]_1^t \\ &= \lim_{t \rightarrow +\infty} \frac{1}{(1-s)t^{s-1}} + \frac{1}{s-1} = \frac{1}{s-1} \end{aligned}$$

si $s > 1$, mientras que si $s < 1$ la integral es divergente. Para $s = 1$ la integral se calcula con la regla del logaritmo y también sale divergente. ■

3.2 Criterios de convergencia

Con las integrales impropias sucede lo mismo que con las series infinitas, que a menudo es posible demostrar su convergencia sin que exista ninguna expresión explícita para el límite (de modo que la propia integral puede usarse para definir números “nuevos”). Muchos criterios de convergencia dependen de la completitud de los números reales a través del teorema siguiente:

Teorema 3.4 (Criterio de Cauchy) Si $f : [a, b[\rightarrow \mathbb{R}$ es una función continua (donde b puede ser $+\infty$) tal que la integral

$$\int_a^b f(x) dx$$

sea impropia en b , entonces es convergente si y sólo si para todo $\epsilon > 0$ existe $a < c < b$ de modo que si $c < u < v < b$ entonces

$$\left| \int_u^v f(x) dx \right| < \epsilon.$$

DEMOSTRACIÓN: Supongamos que se cumple la condición del enunciado y fijemos una sucesión $\{b_n\}_{n=0}^{\infty}$ contenida en $]a, b[$ monótona creciente y convergente a b . Entonces, la sucesión $\{z_n\}_{n=0}^{\infty}$ dada por

$$z_n = \int_a^{b_n} f(x) dx$$

es de Cauchy, pues dado $\epsilon > 0$, tomamos el c que cumple la hipótesis y a su vez tomamos un n_0 tal que si $n \geq n_0$, entonces $c < b_n < b$, con lo que, si $n_0 \leq m \leq n$,

$$|z_n - z_m| = \left| \int_a^{b_n} f(x) dx - \int_a^{b_m} f(x) dx \right| = \left| \int_{b_m}^{b_n} f(x) dx \right| < \epsilon.$$

Por consiguiente, existe $L = \lim_n z_n$.

Ahora, dado $\epsilon > 0$, tomamos $a < c < b$ que cumpla la hipótesis con $\epsilon/2$. Así, si $c < t < b$, existe un n tal que $c < t < b_n < b$ y $|z_n - L| < \epsilon/2$, con lo que

$$\left| \int_a^t f(x) dx - L \right| = \left| \int_a^{b_n} f(x) dx - L - \int_t^{b_n} f(x) dx \right| \leq$$

$$|z_n - L| + \left| \int_t^{b_n} f(x) dx \right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Esto prueba que

$$\lim_{t \rightarrow b} \int_a^t f(x) dx = L,$$

luego la integral es convergente. El recíproco se sigue de la definición de límite y de las propiedades elementales de la integral. ■

Obviamente, se cumple el resultado análogo para integrales impropias en su extremo inferior a . He aquí una primera consecuencia elemental:

Teorema 3.5 Sean $f, g :]a, b[\rightarrow \mathbb{R}$ continuas tales que $0 \leq f(x) \leq g(x)$ para todo $x \in]a, b[$. Si la integral impropia $\int_a^b g(x) dx$ es convergente, lo mismo sucede a la integral de f , y además

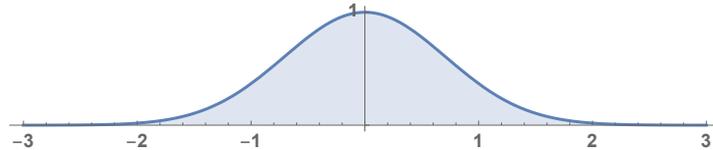
$$\int_a^b f(x) dx \leq \int_a^b g(x) dx.$$

DEMOSTRACIÓN: Es inmediato que el criterio de Cauchy para g implica el criterio de Cauchy para f . ■

Ejemplo La integral

$$\int_{-\infty}^{+\infty} e^{-x^2} dx$$

es convergente.



En efecto, por definición

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \int_{-\infty}^0 e^{-x^2} dx + \int_0^{+\infty} e^{-x^2} dx = 2 \int_0^{+\infty} e^{-x^2} dx,$$

pues el cambio de variable $x = -u$ transforma la primera integral en la segunda.²

²El teorema de cambio de variable es válido para integrales impropias. Por ejemplo, en este caso podemos aplicarlo a

$$\int_t^0 e^{-x^2} dx = - \int_{-t}^0 e^{-u^2} du = \int_0^{-t} e^{-x^2} dx$$

y luego hacer tender t a $+\infty$.

Por lo tanto, basta probar la convergencia de la integral en $[0, +\infty[$, que a su vez equivale a la convergencia en $[1, +\infty[$, pero en este intervalo $1 \leq x$, luego $e^{-x^2} \leq xe^{-x^2}$ y la integral

$$\begin{aligned} \int_1^{+\infty} xe^{-x^2} dx &= \lim_{t \rightarrow +\infty} \int_1^t xe^{-x^2} dx = \lim_{t \rightarrow +\infty} \left[-\frac{1}{2}e^{-x^2} \right]_1^t \\ &= \lim_{t \rightarrow +\infty} \frac{1}{2}e - \frac{1}{2}e^{-t^2} = \frac{e}{2} \end{aligned}$$

es convergente. ■

Otro criterio útil es el siguiente:

Teorema 3.6 (Criterio de Dirichlet) Si $f, g : [a, +\infty[\rightarrow \mathbb{R}$ son funciones continuas de modo que:

a) La primitiva $F(t) = \int_a^t f(x) dx$ está acotada.

b) La función $g(x) \geq 0$ es monótona decreciente, con derivada continua y $\lim_{x \rightarrow +\infty} g(x) = 0$,

entonces la integral

$$\int_a^{+\infty} f(x)g(x) dx$$

es convergente.

DEMOSTRACIÓN: Integramos por partes:

$$\begin{aligned} \int_a^t f(x)g(x) dx &= [g(x)F(x)]_a^t - \int_a^t F(x)g'(x) dx = \\ &= g(t)F(t) - g(a)F(a) - \int_a^t F(x)g'(x) dx. \end{aligned}$$

Por una parte,

$$\lim_{t \rightarrow +\infty} (g(t)F(t) - g(a)F(a)) = -g(a)F(a),$$

donde hemos usado que $g(t)$ tiende a 0 y que $F(t)$ está acotada. Por lo tanto, basta probar que la integral

$$\int_a^{+\infty} F(x)g'(x) dx$$

es convergente, para lo cual empleamos el criterio de Cauchy. Si $a < t_1 < t_2$, entonces

$$\left| \int_{t_1}^{t_2} F(x)g'(x) dx \right| \leq \int_{t_1}^{t_2} |F(x)||g'(x)| dx \leq M \int_{t_1}^{t_2} g'(x) dx,$$

donde M es una cota de F , es decir, que cumple $|F(x)| \leq M$ para todo x . Notemos además que, como $g(x)$ es decreciente, su derivada es $g'(x) \leq 0$. Por lo tanto,

$$\left| \int_{t_1}^{t_2} F(x)g'(x) dx \right| \leq -M(g(t_2) - g(t_1)) = M(g(t_1) - g(t_2)) \leq Mg(t_1)$$

y, como $g(t_1)$ tiende a 0, el módulo de la integral se hace $< \epsilon$ siempre que t_1 y t_2 son suficientemente grandes, como requiere el criterio de Cauchy. ■

Ejemplo Ahora es inmediata la convergencia de las integrales

$$\int_1^{+\infty} \frac{\operatorname{sen} x}{x} dx, \quad \int_1^{+\infty} \frac{\operatorname{cos} x}{x} dx$$

Basta aplicar el criterio de Dirichlet con $g(x) = 1/x$. Notemos que las funciones

$$\int_1^t \operatorname{sen} x dx = \cos 1 - \cos t, \quad \int_1^t \operatorname{cos} x dx = \operatorname{sen} t - \operatorname{sen} 1$$

están acotadas. ■

Otro criterio elemental es el siguiente:

Teorema 3.7 Si $f : [a, +\infty[\rightarrow \mathbb{R}$ es una función continua, $\lim_{x \rightarrow +\infty} f(x) = L$ y la integral

$$\int_a^{+\infty} f(x) dx$$

es convergente, entonces $L = 0$.

DEMOSTRACIÓN: Si $L > 0$, existe un cierto x_0 tal que si $x \geq x_0$ entonces $f(x) \geq L/2$ y, como la integral

$$\int_{x_0}^{+\infty} \frac{L}{2} dx$$

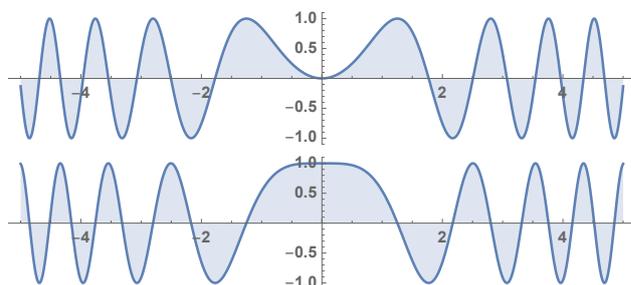
es obviamente divergente, lo mismo le sucede a la de $f(x)$, por 3.5, y la divergencia en $[x_0, +\infty[$ implica claramente la divergencia en $[a, +\infty[$. Si $L < 0$ tenemos que la integral de $-f$ es divergente, luego la de f también. ■

Nota Adaptando mínimamente el argumento del teorema anterior vemos que el integrando de una integral convergente en $[a, +\infty[$ no puede tender a $\pm\infty$ en $+\infty$, pero eso no significa que tenga que converger necesariamente a 0. También cabe la posibilidad de que no tenga límite, como sucede en el ejemplo siguiente:

Ejemplo Las integrales de Fresnel

$$\int_{-\infty}^{+\infty} \operatorname{sen} x^2 dx, \quad \int_{-\infty}^{+\infty} \operatorname{cos} x^2 dx.$$

son convergentes.



En efecto, basta ver que convergen en el intervalo $[1, +\infty[$, para lo cual hacemos el cambio de variable $x = \sqrt{t}$, $dx = dt/2\sqrt{t}$, de modo que

$$\int_1^u \sin x^2 dx = \frac{1}{2} \int_1^{u^2} \frac{\sin t}{\sqrt{t}} dt, \quad \int_1^u \cos x^2 dx = \frac{1}{2} \int_1^{u^2} \frac{\cos t}{\sqrt{t}} dt,$$

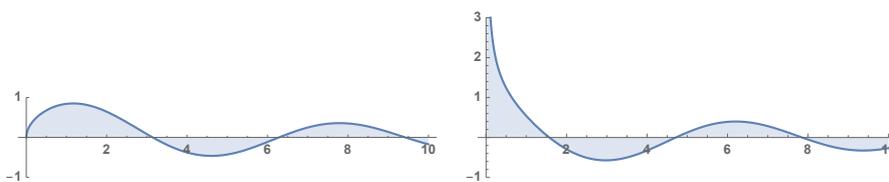
luego

$$\int_1^{+\infty} \sin x^2 dx = \frac{1}{2} \int_1^{+\infty} \frac{\sin t}{\sqrt{t}} dt, \quad \int_1^{+\infty} \cos x^2 dx = \frac{1}{2} \int_1^{+\infty} \frac{\cos t}{\sqrt{t}} dt,$$

y las integrales de los miembros derechos convergen por el criterio de Dirichlet. Conviene observar que el cambio de variable es válido de hecho en todo el intervalo $[0, +\infty[$, de modo que

$$\begin{aligned} \int_{-\infty}^{+\infty} \sin x^2 dx &= 2 \int_0^{+\infty} \sin x^2 dx = \int_0^{+\infty} \frac{\sin u}{\sqrt{u}} du, \\ \int_{-\infty}^{+\infty} \cos x^2 dx &= 2 \int_0^{+\infty} \cos x^2 dx = \int_0^{+\infty} \frac{\cos u}{\sqrt{u}} du, \end{aligned} \quad (3.2)$$

y hemos probado la convergencia en $[0, +\infty[$ de los miembros derechos (lo cual no es inmediato en el caso de la integral del coseno, porque la versión en términos de u es impropia en 0). La figura siguiente muestra las gráficas de los dos integrandos. ■



Veamos un ejemplo de aplicación del criterio del límite.

Ejemplo La integral

$$\int_e^{+\infty} \frac{1}{\log^s x} dx$$

es divergente para todo número real s .

En efecto, si $s < 0$ el integrando tiende a $+\infty$ en $+\infty$, luego la integral diverge. Si $s = 0$ el límite es 1 y la conclusión es la misma, así que basta considerar el caso en que $s > 0$. Hacemos el cambio $x = e^t$, $dx = e^t dt$, con lo que

$$\int_e^u \frac{1}{\log^s x} dx = \int_1^{\log u} \frac{e^t}{t^s} dt \geq \int_1^{\log u} \frac{e^t}{t^n} dt,$$

donde n es cualquier número natural mayor que s , y así, aplicando, por ejemplo, la regla de L'Hôpital, vemos que el integrando tiende a $+\infty$, luego la integral

$$\int_1^{+\infty} \frac{e^t}{t^n} dt$$

es divergente y lo mismo le sucede a la integral del enunciado. ■

Convergencia absoluta Ahora probamos el resultado análogo al hecho de que toda serie absolutamente convergente es convergente:

Teorema 3.8 Si $f :]a, b[\rightarrow \mathbb{R}$ es una función continua y la integral

$$\int_a^b |f(x)| dx$$

es convergente, también lo es la integral sin el valor absoluto, y además

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx.$$

DEMOSTRACIÓN: Vamos a suponer que la integral es impropia únicamente en b . Análogamente se trata el caso en que sólo es impropia en a y el caso en que es impropia en ambos extremos se reduce a éstos dos descomponiendo la integral por un punto intermedio.

Dado $\epsilon > 0$ sea $a < c < b$ tal que la integral de $|f|$ cumpla la condición de Cauchy del teorema anterior. Entonces, si $c < u < v < b$, tenemos que

$$\left| \int_u^v f(x) dx \right| \leq \int_u^v |f(x)| dx < \epsilon,$$

luego f también cumple la condición de Cauchy y su integral en $[a, b]$ es convergente. Como además

$$-\int_a^t |f(x)| dx \leq \int_a^t f(x) dx \leq \int_a^t |f(x)| dx,$$

tomando límites cuando $t \rightarrow b$ obtenemos que

$$-\int_a^b |f(x)| dx \leq \int_a^b f(x) dx \leq \int_a^b |f(x)| dx,$$

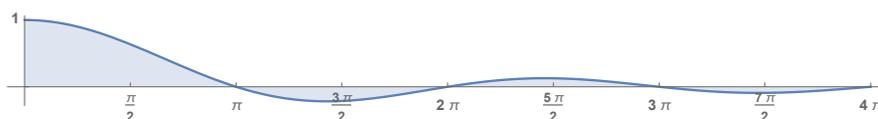
lo cual equivale a la desigualdad del enunciado. ■

Una integral impropia se dice *absolutamente convergente* si la integral del valor absoluto del integrando es convergente. Las integrales que son convergentes, pero no absolutamente convergentes, se dicen *condicionalmente convergentes*.

Ejemplo La *integral de Dirichlet*

$$\int_0^{+\infty} \frac{\operatorname{sen} x}{x} dx$$

es convergente, pero no absolutamente convergente.



Tras el teorema 3.6 hemos visto que la integral es convergente (el integrando se extiende a una función continua en $[0, 1]$, por lo que la convergencia en el intervalo $[0, +\infty[$ equivale a la convergencia en $[1, +\infty[$). Como $|\operatorname{sen} x| \leq 1$, se cumple que $\operatorname{sen}^2 x = |\operatorname{sen} x|^2 \leq |\operatorname{sen} x|$, luego

$$\frac{\operatorname{sen}^2 x}{x} \leq \frac{|\operatorname{sen} x|}{x}.$$

Vamos a probar que la integral

$$\int_1^{+\infty} \frac{\operatorname{sen}^2 x}{x} dx$$

es divergente, con lo que el teorema 3.5 implica que

$$\int_1^{+\infty} \frac{|\operatorname{sen} x|}{x} dx$$

también lo es, luego la integral del enunciado no es absolutamente convergente.

Ahora bien:

$$\begin{aligned} \int_1^t \frac{\operatorname{sen}^2 x}{x} dx &= \frac{1}{2} \int_1^t \frac{1 - \cos 2x}{x} dx = \frac{1}{2} \int_1^t \frac{1}{x} dx - \int_1^t \frac{\cos 2x}{2x} dx \\ &= \frac{\log t}{2} - \frac{1}{2} \int_1^t \frac{\cos x}{x} dx, \end{aligned}$$

y hemos visto que la segunda integral es convergente, luego

$$\lim_{t \rightarrow +\infty} \int_1^t \frac{\operatorname{sen}^2 x}{x} dx = +\infty. \quad \blacksquare$$

Nota Es fácil ver que si llamamos

$$a_n = \int_{n\pi}^{(n+1)\pi} \frac{\operatorname{sen} x}{x} dx,$$

entonces

$$\int_0^{+\infty} \frac{\operatorname{sen} x}{x} dx = \sum_{n=0}^{\infty} a_n,$$

y de hecho la convergencia de la serie (que es una serie alternada con término general convergente a 0, luego converge por el criterio de Leibniz [ITAn 2.9]) implica la convergencia de la integral. Que ésta no sea absolutamente convergente equivale a su vez a que la serie tampoco lo es. ■

El mismo argumento con el que hemos probado en el ejemplo precedente que la convergencia no es absoluta puede usarse para probar que tampoco lo es la de las integrales de Fresnel. Veamos otro ejemplo del mismo estilo:

Ejemplo *La integral*

$$\int_e^{+\infty} \frac{\operatorname{sen} x}{\log^s x} dx$$

es convergente, pero no absolutamente convergente, para todo $s > 0$.

En efecto, la convergencia la garantiza el criterio de Dirichlet y, para probar que no es absolutamente convergente, tenemos en cuenta que

$$\frac{1}{2} \left(\frac{1}{\log^s x} - \frac{\cos 2x}{\log^s x} \right) = \frac{\operatorname{sen}^2 x}{\log^s x} \leq \frac{|\operatorname{sen} x|}{\log^s x},$$

luego basta probar que la integral de la función de la izquierda es divergente, pero la del segundo término es convergente por el criterio de Dirichlet, y en el ejemplo de la página 155 hemos visto que la del primero es divergente, luego la integral completa es divergente. ■

Terminamos esta sección generalizando los teoremas 2.8 y 2.16:

Teorema 3.9 (Teorema de la convergencia monótona) *Sea $\{f_n\}_{n=0}^{\infty}$ una sucesión de funciones continuas $f_n :]a, b[\rightarrow \mathbb{R}$ y sea $f :]a, b[\rightarrow \mathbb{R}$ otra función, también continua, de modo que, para todo $x \in]a, b[$ y todos los números naturales $m \leq n$, se cumpla que $f_m(x) \leq f_n(x) \leq f(x)$ y $\lim_n f_n(x) = f(x)$. Si las integrales*

$$\int_a^b f_n(x) dx, \quad \int_a^b f(x) dx$$

son convergentes, se cumple que

$$\int_a^b f(x) dx = \lim_n \int_a^b f_n(x) dx.$$

DEMOSTRACIÓN: Llamamos $g_n(x) = f(x) - f_n(x)$, de modo que, para todo $x \in]a, b[$ y $m \leq n$, se cumple que $0 \leq g_n(x) \leq g_m(x)$ y $\lim_n g_n(x) = 0$. Tenemos que probar que la sucesión de integrales converge a 0.

Dado $\epsilon > 0$, existen $a < a' < b' < b$ tales que

$$\int_a^{a'} g_0(x) dx + \int_{b'}^b g_0(x) dx < \frac{\epsilon}{2},$$

luego también

$$\int_a^{a'} g_n(x) dx + \int_{b'}^b g_n(x) dx < \frac{\epsilon}{2},$$

para todo n . Por 2.8 sabemos que

$$\lim_n \int_{a'}^{b'} f_n(x) dx = \int_{a'}^{b'} f(x) dx$$

o, equivalentemente.

$$\lim_n \int_{a'}^{b'} g_n(x) dx = 0.$$

Por lo tanto, existe un n_0 tal que si $n \geq n_0$, se cumple que

$$\int_{a'}^{b'} g_n(x) dx < \frac{\epsilon}{2},$$

luego

$$\int_a^b g_n(x) dx < \epsilon. \quad \blacksquare$$

Teorema 3.10 (De la convergencia dominada) *Sea $\{f_n\}_{n=0}^\infty$ una sucesión de funciones continuas $f_n :]a, b[\rightarrow \mathbb{R}$ (donde a y b pueden ser infinitos) tales que existen funciones continuas $\phi, f :]a, b[\rightarrow \mathbb{R}$ con integral convergente y tales que, para todo $x \in]a, b[$ se cumple que*

$$|f_n(x)| \leq \phi(x), \quad \lim_n f_n(x) = f(x).$$

Entonces

$$\lim_n \int_a^b f_n(x) dx = \int_a^b f(x) dx.$$

DEMOSTRACIÓN: Notemos que las hipótesis implican que $|f(x)| \leq \phi(x)$, luego la convergencia de la integral de f se sigue en realidad de la de ϕ . Dado $\epsilon > 0$, existen $a < a' < b' < b$ tales que

$$\int_a^{a'} \phi(x) dx + \int_{b'}^b \phi(x) dx < \frac{\epsilon}{4}, \quad \left| \int_a^{a'} f(x) dx \right| + \left| \int_{b'}^b f(x) dx \right| < \frac{\epsilon}{4}.$$

Las funciones f_n restringidas al intervalo $[a', b']$ están acotadas por una cota de $\phi(x)$, luego el teorema 2.16 implica que

$$\lim_n \int_{a'}^{b'} f_n(x) dx = \int_{a'}^{b'} f(x) dx,$$

luego existe un n_0 tal que si $n \geq n_0$, se cumple que

$$\left| \int_{a'}^{b'} f_n(x) dx - \int_{a'}^{b'} f(x) dx \right| < \frac{\epsilon}{2},$$

luego

$$\left| \int_a^b f_n(x) dx - \int_a^b f(x) dx \right| \leq \left| \int_{a'}^{b'} f_n(x) dx - \int_{a'}^{b'} f(x) dx \right| + \int_a^{a'} \phi(x) dx + \int_{b'}^b \phi(x) dx + \left| \int_a^{a'} f(x) dx \right| + \left| \int_{b'}^b f(x) dx \right| < \epsilon. \quad \blacksquare$$

3.3 Derivación de integrales paramétricas

Consideremos una función de dos variables, por ejemplo $f(x, t) = \cos(t^2x)$. Ésta determina dos familias de funciones de una variable, a saber, las funciones $f_x(t) = \cos(t^2x)$ y $f_t(x) = \cos(t^2x)$, y —en este caso concreto— vemos que cualquiera de las funciones de cualquiera de las dos familias es derivable. Las derivadas de estas funciones se llaman *derivadas parciales* de la función f , y se representan así:

$$\frac{\partial f}{\partial x} = \frac{df_t}{dx} = -t^2 \operatorname{sen}(t^2x), \quad \frac{\partial f}{\partial t} = \frac{df_x}{dt} = -2tx \operatorname{sen}(t^2x).$$

Estas derivadas parciales las estudiaremos con detalle en el capítulo V, pero de momento no necesitamos más que su mera definición. Podemos pensar que $\partial f / \partial t$ es la derivada de la función que resulta de considerar a f como una función de f en la que t es un parámetro. En esta sección nos vamos a ocupar de las funciones definidas mediante integrales paramétricas, como

$$f(t) = \int_0^1 \cos(t^2x) dx = \frac{1}{t^2} [\operatorname{sen}(t^2x)]_0^1 = \frac{\operatorname{sen} t^2}{t^2}.$$

Más precisamente,

$$f(t) = \begin{cases} \frac{\operatorname{sen} t^2}{t^2} & \text{si } t \neq 0, \\ 1 & \text{si } t = 0. \end{cases}$$

No es difícil comprobar que se cumple la relación siguiente:

$$\frac{d}{dt} \int_0^1 \cos(t^2x) dx = \int_0^1 \frac{\partial \cos(t^2x)}{\partial t} dx.$$

Explícitamente, para $t \neq 0$ es

$$\frac{2t^3 \cos t^2 - 2t \operatorname{sen} t^2}{t^4} = -2t \int_0^1 x \operatorname{sen}(t^2x) dx$$

y para $t = 0$ es $f'(0) = 0$.

El resultado principal de esta sección es que, bajo hipótesis razonables, esto es cierto en general, lo cual, como veremos, tiene muchas consecuencias de interés.

Teorema 3.11 Sea $f :]a, b[\times]c, d[\rightarrow \mathbb{R}$ una función continua, donde a y c pueden ser $-\infty$ y b y d pueden ser $+\infty$, de modo que, para todo $t \in]c, d[$, exista la integral

$$g(t) = \int_a^b f(x, t) dx.$$

Supongamos además que existe

$$\frac{\partial f}{\partial t} :]a, b[\times]c, d[\rightarrow \mathbb{R}$$

y que es continua, así como que existe una función continua $\phi :]a, b[\rightarrow \mathbb{R}$ de modo que la integral

$$\int_a^b \phi(x) dx$$

es convergente y

$$\left| \frac{\partial f}{\partial t} \right| \leq \phi(x)$$

para todo $(x, t) \in]a, b[\times]c, d[$. Entonces la función $g(t)$ es derivable y

$$\frac{d}{dt} \int_a^b f(x, t) dx = \int_a^b \frac{\partial f}{\partial t} dx.$$

DEMOSTRACIÓN: Tomamos $a < e < b$ y partimos la integral

$$g(t) = g_1(t) + g_2(t) = \int_a^e f(x, t) dx + \int_e^b f(x, t) dx.$$

Basta probar el teorema para g_1 y g_2 . Consideramos el caso de g_2 (el de g_1 es análogo) y vemos que, equivalentemente, llamando a a e , podemos suponer que

$$f :]a, b[\times]c, d[\rightarrow \mathbb{R}, \quad \phi :]a, b[\rightarrow \mathbb{R}$$

son continuas. Vamos a ver que $g(t)$ es derivable en un punto concreto t , para lo cual, restringiendo el intervalo $]c, d[$ a un intervalo $[t - \eta, t + \eta] \subset]c, d[$, podemos suponer que

$$f :]a, b[\times [c, d] \rightarrow \mathbb{R}$$

es continua.

Para cada h tal que $t + h \in]c, d[$, tenemos que

$$\left| \frac{g(t+h) - g(t)}{h} - \int_a^b \frac{\partial f}{\partial t} dx \right| = \left| \int_a^b \left(\frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right) dx \right| \leq$$

$$\left| \int_a^{b^*} \left(\frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right) dx \right| + \left| \int_{b^*}^b \left(\frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right) dx \right|,$$

donde, dado $\epsilon > 0$, elegimos b^* de modo que

$$\int_{b^*}^b \phi(x) dt < \frac{\epsilon}{3}.$$

Fijado un valor de x , el teorema del valor medio nos da un \tilde{t} entre t y $t+h$ tal que

$$f(x, t+h) - f(x, t) = \left. \frac{\partial f}{\partial t} \right|_{(x, \tilde{t})} h,$$

por lo que

$$\left| \frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right| = \left| \left. \frac{\partial f}{\partial t} \right|_{(x, \tilde{t})} - \left. \frac{\partial f}{\partial t} \right|_{(x, t)} \right| \leq \left| \left. \frac{\partial f}{\partial t} \right|_{(x, \tilde{t})} \right| + \left| \left. \frac{\partial f}{\partial t} \right|_{(x, t)} \right| \leq 2\phi(x).$$

El teorema 3.5 nos da que el primer término de la cadena de desigualdades es una función integrable en $[b^*, b]$, y 3.8 implica entonces que

$$\begin{aligned} \int_{b^*}^b \left(\frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right) dt &\leq \int_{b^*}^b \left| \frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right| dt \leq \\ &2 \int_{b^*}^b \phi(x) dx < \frac{2\epsilon}{3}. \end{aligned}$$

Por otro lado, como f es continua en el compacto $K = [a, b^*] \times [c, d]$, el teorema de Heine-Cantor [ITAn .14] nos da un $\delta > 0$ de manera que si $(x_1, t_1), (x_2, t_2) \in K$ cumplen $\|(x_1, t_1) - (x_2, t_2)\| < \delta$, entonces

$$\left| \left. \frac{\partial f}{\partial t} \right|_{(x_1, t_1)} - \left. \frac{\partial f}{\partial t} \right|_{(x_2, t_2)} \right| < \frac{\epsilon}{3(b^* - a)}.$$

En particular, si $|h| < \delta$ y $x \in [a, b^*]$, se cumple que

$$\left| \frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right| = \left| \left. \frac{\partial f}{\partial t} \right|_{(x, \tilde{t})} - \left. \frac{\partial f}{\partial t} \right|_{(x, t)} \right| < \frac{\epsilon}{3(b^* - a)},$$

ya que $\|(x, \tilde{t}) - (x, t)\| \leq |\tilde{t} - t| < h < \delta$. Por consiguiente,

$$\begin{aligned} \left| \int_a^{b^*} \left(\frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right) dx \right| &\leq \int_a^{b^*} \left| \frac{f(x, t+h) - f(x, t)}{h} - \frac{\partial f}{\partial t} \right| dx \\ &\leq \int_a^{b^*} \frac{\epsilon}{3(b^* - a)} dt = \frac{\epsilon}{3}. \end{aligned}$$

En definitiva, dado $\epsilon > 0$, hemos encontrado un $\delta > 0$, tal que, si $|h| < \delta$,

$$\left| \frac{g(t+h) - g(t)}{h} - \int_a^b \frac{\partial f}{\partial t} dx \right| < \epsilon.$$

Esto prueba la existencia de

$$\frac{dg}{dt} = \lim_{h \rightarrow 0} \frac{g(t+h) - g(t)}{h} = \int_a^b \frac{\partial f}{\partial t} dt. \quad \blacksquare$$

Nota Si en el teorema anterior partimos de una función $f : [a, b] \times]c, d[\rightarrow \mathbb{R}$ continua y cuya derivada admite una extensión continua:

$$\frac{\partial f}{\partial t} : [a, b] \times]c, d[\rightarrow \mathbb{R},$$

entonces podemos eliminar la hipótesis sobre la función ϕ tomando en la demostración $b^* = b$, es decir, no hace falta considerar las integrales como impropias. Equivalentemente, para probar la relación en un punto x_0 , podemos restringir f a $K = [a, b] \times [x_0 - \delta, x_0 + \delta]$ y tomar $\phi(x) = C$, donde C es una cota de la función continua $\partial f / \partial x$ en compacto K . \blacksquare

Este teorema puede usarse para calcular integrales definidas de funciones de las que no podemos calcular una primitiva.

Ejemplo

$$\int_0^1 \frac{\log(x+1)}{1+x^2} dx = \frac{\pi \log 2}{8}.$$

En efecto, para calcular la integral consideramos la función

$$g(t) = \int_0^1 \frac{\log(tx+1)}{1+x^2} dx,$$

de modo que la integral que queremos calcular es $g(1)$. Notemos que el integrando es continuo, digamos para $(x, t) \in [0, 1] \times]0, 2[$, por lo que la integral no es impropia. Por el teorema anterior tenemos que

$$g'(t) = \int_0^1 \frac{x}{(tx+1)(1+x^2)} dx = \frac{1}{1+t^2} \int_0^1 \left(\frac{x+t}{1+x^2} - \frac{t}{tx+1} \right) dx,$$

donde hemos calculado la descomposición en fracciones simples del integrando. Así:

$$\begin{aligned} g'(t) &= \frac{1}{1+t^2} \left([t \arctan x]_0^1 + \frac{1}{2} [\log(1+x^2)]_0^1 - [\log(tx+1)]_0^1 \right) \\ &= \frac{1}{1+t^2} \left(\frac{t\pi}{4} + \frac{\log 2}{2} - \log(t+1) \right) = \frac{\pi t + 2 \log 2 - 4 \log(t+1)}{4(1+t^2)}. \end{aligned}$$

Puesto que $g(0) = 0$, tenemos que

$$\begin{aligned} g(1) &= \int_0^1 g'(t) dt = \int_0^1 \frac{\pi t + 2 \log 2 - 4 \log(t+1)}{4(1+t^2)} dt \\ &= \int_0^1 \frac{\pi t + 2 \log 2}{4(1+t^2)} dt - g(1), \end{aligned}$$

luego

$$\begin{aligned} g(1) &= \frac{1}{8} \int_0^1 \frac{\pi t + 2 \log 2}{1+t^2} dt = \frac{1}{8} \left[\frac{\pi}{2} \log(1+t^2) + 2 \log 2 \arctan t \right]_0^1 \\ &= \frac{1}{8} \left(\frac{\pi}{2} \log 2 + 2 \log 2 \frac{\pi}{4} \right) = \frac{\pi \log 2}{8}. \end{aligned}$$

■

Veamos ahora un ejemplo con una integral impropia:

Ejemplo: La integral de Gauss

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}.$$

En la página 152 hemos visto que la integral es convergente, así como que

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = 2 \int_0^{+\infty} e^{-x^2} dx.$$

Por lo tanto, basta calcular la segunda integral, a la que llamaremos I . Para ello consideramos la función

$$g(t) = \int_0^{+\infty} \frac{e^{-t^2(1+x^2)}}{1+x^2} dx.$$

Notemos que, para cada t , la integral converge por el criterio de Dirichlet, tomando como función decreciente el numerador. El teorema 3.11 nos da que

$$g'(t) = - \int_0^{+\infty} 2te^{-t^2(1+x^2)} dx.$$

Para justificar que esta igualdad es válida para un $t > 0$ arbitrario tomamos $0 < \delta < t < c$ y aplicamos el teorema 3.11 considerando el integrando definido en $]0, +\infty[\times]\delta, c[$ y la función

$$\phi(x) = 2ce^{-\delta^2(1+x^2)} = 2ce^{-\delta^2} e^{-(\delta x)^2},$$

que es integrable en $]0, +\infty[$ porque se transforma en e^{-x^2} (salvo un factor constante) con el cambio $\delta x \mapsto x$.

Así pues, haciendo el cambio $y = tx$,

$$g'(t) = -2te^{-t^2} \int_0^{+\infty} e^{-(tx)^2} dx = -2e^{-t^2} \int_0^{+\infty} e^{-y^2} dy = -2Ie^{-t^2}.$$

Ahora observamos que

$$g(0) = \int_0^{+\infty} \frac{dx}{1+x^2} = \lim_{t \rightarrow +\infty} [\arctan x]_0^t = \frac{\pi}{2},$$

luego

$$g(t) = \frac{\pi}{2} + \int_0^t g'(u) du = \frac{\pi}{2} - 2I \int_0^t e^{-t^2} dt,$$

luego

$$\lim_{t \rightarrow +\infty} g(t) = \frac{\pi}{2} - 2I^2.$$

Sólo falta probar que el límite vale 0. Para ello observamos que, para $t \geq 1$,

$$g(t) = \int_0^{+\infty} \frac{e^{-t^2(1+x^2)}}{1+x^2} dx \leq e^{-t^2} \int_0^{+\infty} e^{-t^2 x^2} dx \leq e^{-t^2} \int_0^{+\infty} e^{-x^2} dx,$$

luego, en efecto, el límite es 0. ■

Veamos un ejemplo que muestra que la conclusión del teorema 3.11 puede fallar si no se cumplen sus hipótesis:

Ejemplo Consideremos la función

$$g(t) = \int_0^1 \frac{t^3}{x^2} e^{-t^2/x} dx = \lim_{u \rightarrow 0^+} \int_u^1 \frac{t^3}{x^2} e^{-t^2/x} dx =$$

$$\lim_{u \rightarrow 0^+} [te^{-t^2/x}]_u^1 = \lim_{u \rightarrow 0^+} te^{-t^2} - te^{-t^2/u} = te^{-t^2}.$$

Tenemos que

$$\frac{d}{dt} \int_0^1 \frac{t^3}{x^2} e^{-t^2/x} dx = e^{-t^2} (1 - 2t^2).$$

Por otra parte,

$$\int_0^1 \frac{\partial}{\partial t} \left(\frac{t^3}{x^2} e^{-t^2/x} \right) dx = \int_0^1 e^{-t^2/x} \left(\frac{3t^2}{x^2} - \frac{2t^4}{x^3} \right) dx$$

Si $t \neq 0$, integrando por partes el segundo sumando, queda que

$$\int e^{-t^2/x} \left(\frac{3t^2}{x^2} - \frac{2t^4}{x^3} \right) dx = e^{-t^2/x} \left(1 - \frac{2t^2}{x} \right) + c,$$

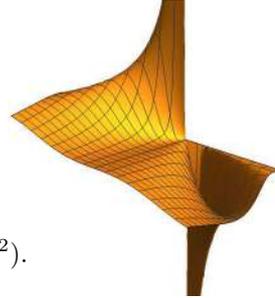
de donde

$$\int_0^1 \frac{\partial}{\partial t} \left(\frac{t^3}{x^2} e^{-t^2/x} \right) dx = e^{-t^2} (1 - 2t^2) = g'(t).$$

Sin embargo, este cálculo no vale para $t = 0$, en cuyo caso el integrando es nulo, luego la integral también. Así pues, $g'(0) = 1$, que no coincide con la integral de la derivada. ■

En la sección siguiente veremos más aplicaciones de la derivación de integrales paramétricas. Terminamos enunciando una generalización del teorema 3.11 para integrales propias cuya demostración posponemos hasta 5.13, pues entonces será inmediata.

$$\frac{d}{dx} \int_{a(x)}^{b(x)} f(x, t) dt = f(x, b(x))b'(x) - f(x, a(x))a'(x) + \int_{a(x)}^{b(x)} \frac{\partial f}{\partial x} dt. \quad (3.3)$$



3.4 Integrales dobles

Muchos argumentos en los que intervienen series infinitas tienen como paso clave el intercambio de dos sumatorios. También hay argumentos análogos que se basan en el intercambio de una suma y una integral, o en el intercambio de dos integrales. Vamos a ocuparnos ahora de este último caso, aunque sólo veremos algunos casos particulares que vamos a necesitar. El caso más simple es el siguiente:

Teorema 3.12 Si $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$ es una función continua, entonces

$$\int_a^b \left(\int_c^d f(x, y) dx \right) dy = \int_c^d \left(\int_a^b f(x, y) dy \right) dx.$$

DEMOSTRACIÓN: Llamemos

$$F(x, y) = \int_a^x f(t, y) dt.$$

Como función de x es una primitiva de $f(x, y)$, es decir, que se cumple que

$$\frac{\partial F}{\partial x} = f(x, y).$$

Por el teorema 3.11 (sin necesidad de considerar ninguna función ϕ , pues todas las integrales son propias), tenemos que

$$\frac{d}{dx} \int_c^d \left(\int_a^x f(t, y) dt \right) dy = \int_c^d f(x, y) dy,$$

y por otra parte

$$\frac{d}{dx} \int_a^x \left(\int_c^d f(t, y) dy \right) dx = \int_c^d f(x, y) dy.$$

Como las dos funciones tienen la misma derivada, según 1.16 existe una constante k tal que, para todo x en el intervalo $[a, b]$,

$$\int_c^d \left(\int_a^x f(t, y) dt \right) dy = \int_a^x \left(\int_c^d f(t, y) dy \right) dx + k.$$

Pero tomando $x = a$ concluimos que $k = 0$, y haciendo $x = b$ obtenemos la igualdad del enunciado. ■

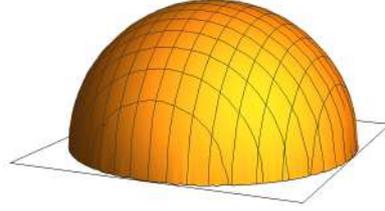
Interpretación geométrica de las integrales dobles Las integrales dobles tienen una interpretación geométrica natural análoga a la de las integrales “simples”. No vamos a justificarla porque el cálculo integral que hemos expuesto es demasiado rudimentario como para sacarle partido, pero la ilustraremos con un ejemplo.

Consideremos la función $f : [-r, r] \times [-r, r] \rightarrow \mathbb{R}$ dada por

$$f(x, y) = \begin{cases} \sqrt{r^2 - x^2 - y^2} & \text{si } x^2 + y^2 \leq r^2, \\ 0 & \text{en otro caso} \end{cases}$$

y vamos a calcular

$$\int_{-r}^r \int_{-r}^r f(x, y) \, dy \, dx.$$



Para ello observamos que la función $f_x(y) = f(x, y)$ toma el valor 0 fuera del intervalo $[-\sqrt{r^2 - x^2}, \sqrt{r^2 - x^2}]$, por lo que su integral en $[-r, r]$ coincide con su integral en este intervalo, luego, usando (C.7):

$$\begin{aligned} \int_{-r}^r \int_{-r}^r f(x, y) \, dy \, dx &= \int_{-r}^r \int_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} \sqrt{r^2 - x^2 - y^2} \, dy \, dx = \\ \int_{-r}^r \left[\frac{y}{2} \sqrt{r^2 - x^2 - y^2} - \frac{r^2 - x^2}{2} \arccos \frac{y}{\sqrt{r^2 - x^2}} \right]_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} dx &= \\ \frac{\pi}{2} \int_{-r}^r (r^2 - x^2) \, dx &= \frac{\pi}{2} \left[r^2 x - \frac{x^3}{3} \right]_{-r}^r = \frac{2}{3} \pi r^3. \end{aligned}$$

El resultado es claramente el volumen de una semiesfera de radio r , es decir, el volumen de la región situada entre el plano XY y la gráfica de f . Esto no es casual, sino que es precisamente la interpretación general de la integral doble de una función de dos variables: es el volumen de la región comprendida entre el plano XY y la gráfica de la función, entendiendo que dicho volumen es negativo en las zonas donde la gráfica está por debajo del plano. Más precisamente, la integral de la función $f_x(y)$ nos da el área de la sección de dicha región por el plano vertical formado por los puntos de segunda coordenada igual a x , que en este caso es un semicírculo de radio $\sqrt{r^2 - x^2}$, por lo que la integral nos ha dado $\pi(r^2 - x^2)$, y así el volumen de la semiesfera se obtiene integrando las áreas de los semicírculos que determina mediante cortes con un haz de planos paralelos. (La figura muestra algunos de estos semicírculos.)

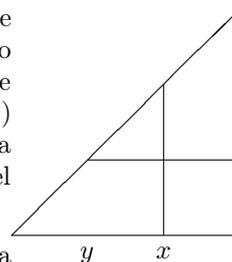
Tal y como hemos indicado, no vamos a justificar este hecho en general, sino que usaremos las integrales dobles como meros instrumentos de cálculo. ■

Hemos probado que dos integrales sobre un rectángulo se pueden calcular en cualquier orden, pero vamos a necesitar una variante para el caso de un triángulo. Con una teoría de integración más fina que permitiera integrar funciones discontinuas razonables, el teorema siguiente sería un caso particular de 3.12:

Teorema 3.13 Consideremos el triángulo $T = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq x \leq b\}$ y sea $f : T \rightarrow \mathbb{R}$ una función continua. Entonces

$$\int_0^b \int_0^x f(x, y) \, dy \, dx = \int_0^b \int_y^b f(x, y) \, dx \, dy.$$

En la integral de la izquierda fijamos un valor de x e integramos $f_x(y)$ desde 0 hasta x (es decir, sobre el segmento vertical que muestra la figura), mientras que en la integral de la derecha fijamos un valor de y e integramos la función $f_y(x)$ desde y hasta b (sobre el segmento horizontal que muestra la figura). Por lo tanto, en ambos casos “barremos” todo el triángulo T , sea con segmentos verticales u horizontales.



DEMOSTRACIÓN: La prueba es idéntica a la del teorema anterior:

$$\frac{d}{du} \int_0^u \int_0^x f(x, y) dy dx = \int_0^u f(u, y) dy.$$

Para derivar el segundo miembro usamos (3.3), con $a(u) = u$, $b(u) = 0$, y donde la función $f(x, t)$ del teorema es

$$F(u, y) = \int_y^u f(x, y) dx,$$

de modo que

$$\frac{\partial F}{\partial u} = f(u, y).$$

Entonces:

$$\frac{d}{du} \int_0^u \int_y^u f(x, y) dx dy = \int_u^u f(x, u) dx + \int_0^u f(u, y) dy = \int_0^u f(u, y) dy.$$

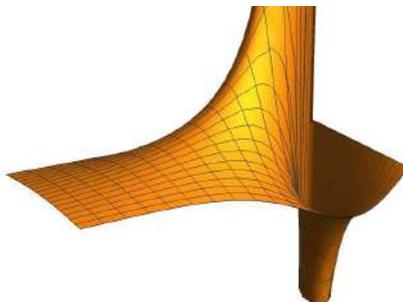
Como las dos funciones tienen la misma derivada, existe una constante k tal que

$$\int_0^u \int_0^x f(x, y) dy dx = \int_0^u \int_y^u f(x, y) dx dy + k,$$

pero haciendo $u = 0$ vemos que $k = 0$, luego haciendo $u = b$ tenemos la fórmula del enunciado. ■

Ejemplo Vamos a comprobar que

$$\int_0^2 \int_0^1 \frac{xy(x^2 - y^2)}{(x^2 + y^2)^3} dy dx = \frac{1}{5}, \quad \int_0^1 \int_0^2 \frac{xy(x^2 - y^2)}{(x^2 + y^2)^3} dx dy = -\frac{1}{20}.$$



Si $x \neq 0$ hacemos el cambio de variable $u = x^2 + y^2$, $du = 2y dy$:

$$\begin{aligned} \int_0^1 \frac{xy(x^2 - y^2)}{(x^2 + y^2)^3} dy &= \frac{1}{2} \int_{x^2}^{x^2+1} \frac{x(2x^2 - u)}{u^3} du = \int_{x^2}^{x^2+1} \left(\frac{x^3}{u^3} - \frac{x}{2u^2} \right) du = \\ &= \left[-\frac{x^3}{2u^2} + \frac{x}{2u} \right]_{x^2}^{x^2+1} = -\frac{x^3}{2(x^2+1)^2} + \frac{x}{2(x^2+1)} + \frac{1}{2x} - \frac{1}{2x} = \frac{x}{2(x^2+1)^2}. \end{aligned}$$

Notemos que la igualdad vale también para $x = 0$, pues en tal caso el integrando es nulo. A su vez,

$$\int_0^2 \frac{x}{2(x^2+1)^2} dx = - \left[\frac{1}{4(x^2+1)} \right]_0^2 = -\frac{1}{20} + \frac{1}{4} = \frac{1}{5}.$$

Por otra parte, haciendo $u = x^2 + y^2$, $du = 2x dx$:

$$\begin{aligned} \int_0^2 \frac{xy(x^2 - y^2)}{(x^2 + y^2)^3} dx &= \frac{1}{2} \int_{y^2}^{4+y^2} \frac{y(u - 2y^2)}{u^3} du = \int_{y^2}^{4+y^2} \left(\frac{y}{2u^2} - \frac{y^3}{u^3} \right) du = \\ &= \left[-\frac{y}{2u} + \frac{y^3}{2u^2} \right]_{y^2}^{4+y^2} = -\frac{y}{2(4+y^2)} + \frac{y^3}{2(4+y^2)^2} + \frac{1}{2y} - \frac{1}{2y} = \frac{-2y}{(4+y^2)^2}. \end{aligned}$$

Como antes, la fórmula vale igualmente para $y = 0$. Finalmente:

$$\int_0^1 \frac{-2y}{(4+y^2)^2} dy = \left[\frac{1}{4+y^2} \right]_0^1 = \frac{1}{5} - \frac{1}{4} = -\frac{1}{20}.$$

La razón por la que el teorema 3.12 no es aplicable a este ejemplo es que el integrando no es continuo en $[0, 2] \times [0, 1]$. La única excepción es el punto $(0, 0)$, pero es suficiente. Las integrales dobles no pueden interpretarse como el volumen limitado por la gráfica de la función pues esta limita una zona de volumen infinito sobre el plano XY otra, también de volumen infinito, bajo dicho plano, luego no hay ningún volumen definido. Las integrales resultan finitas porque en cada integral interior se compensa una parte positiva con otra negativa, de modo que el promedio permanece siempre acotado. ■

Para que una función no acotada pueda definir una integral doble independiente del orden de integración basta con que las integrales iteradas sean absolutamente convergentes en el sentido siguiente:

Teorema 3.14 Sea $f :]a, b[\times]c, d[\rightarrow \mathbb{R}$ una función continua (donde los extremos de los intervalos pueden ser infinitos) tal que las integrales

$$\int_a^b |f(x, y)| dx, \quad y \quad \int_c^d \int_a^b |f(x, y)| dx dy$$

sean convergentes. Entonces

$$\int_a^b \int_c^d f(x, y) dy dx = \int_c^d \int_a^b f(x, y) dx dy.$$

En lugar de demostrar este teorema demostraremos la versión para un triángulo que necesitaremos en la sección siguiente. La prueba del teorema anterior es ligeramente más simple.

Teorema 3.15 Consideremos el triángulo infinito $T = \{(x, y) \in \mathbb{R} \mid 0 \leq y \leq x\}$ y sea $f : T \rightarrow \mathbb{R}$ una función continua. Si las integrales

$$\int_y^{+\infty} |f(x, y)| dx, \quad \int_0^{+\infty} \int_y^{+\infty} |f(x, y)| dx dy$$

son convergentes, entonces

$$\int_0^{+\infty} \int_0^x f(x, y) dy dx = \int_0^{+\infty} \int_y^{+\infty} f(x, y) dx dy.$$

DEMOSTRACIÓN: Partimos de la igualdad

$$\int_0^b \int_0^x f(x, y) dy dx = \int_0^b \int_y^b f(x, y) dx dy$$

que nos da el teorema anterior para todo $b > 0$. Para cada número natural n definimos

$$h_n(y) = \begin{cases} \int_y^n f(x, y) dx & \text{si } y \leq n, \\ 0 & \text{si } n \leq y. \end{cases}$$

En estos términos, si $n \leq m$, la igualdad anterior, para $b = n$, se puede expresar así:

$$\int_0^n \int_0^x f(x, y) dy dx = \int_0^m h_n(y) dy.$$

Como el miembro izquierdo no depende de m , trivialmente

$$\int_0^n \int_0^x f(x, y) dy dx = \int_0^{+\infty} h_n(y) dy.$$

Observemos ahora que $h_n : [0, +\infty[\rightarrow \mathbb{R}$ es una función continua. En efecto, la integral es $F(n) - F(y)$, donde F es una primitiva de $f_y(x)$, luego es una función derivable y, en particular, continua, y toma el valor 0 cuando $y = n$, por lo que la función h_n es continua en n , y trivialmente también en los puntos $y > n$. Además, $\lim_n h_n(y) = h(y)$, donde

$$h(y) = \int_y^{+\infty} f(x, y) dx,$$

que es una de las integrales del miembro derecho del enunciado que suponemos (absolutamente) convergentes. Además,

$$|h_n(y)| \leq \int_y^n |f(x, y)| dx \leq \int_y^{+\infty} |f(x, y)| dx = \phi(y)$$

y estamos suponiendo que $\int_0^{+\infty} \phi(y) dy$ es convergente, luego se cumplen las hipótesis del teorema de la convergencia dominada 3.10, que nos permite concluir³ que

$$\lim_n \int_0^{+\infty} h_n(y) dy = \int_0^{+\infty} h(y) dy = \int_0^{+\infty} \int_y^{+\infty} f(x, y) dx dy,$$

y esto equivale a la convergencia de

$$\int_0^{+\infty} \int_0^x f(x, y) dy dx = \int_0^{+\infty} \int_y^{+\infty} f(x, y) dx dy. \quad \blacksquare$$

3.5 La transformada de Laplace

Veamos ahora cómo las integrales impropias nos proporcionan una herramienta muy potente que, entre otras cosas, puede transformar ciertas ecuaciones diferenciales en ecuaciones algebraicas. Lo que ahora se conoce como “transformada de Laplace” es un concepto que aparece implícito en algunos resultados de Euler, Lagrange, Abel, y Laplace, aunque no fue hasta principios del siglo XX cuando el ingeniero Olivier Heaviside lo desarrolló en forma de lo que se llamó el “cálculo operacional de Heaviside” consistente en sustituir las derivadas por un “operador p ” en las ecuaciones diferenciales, el cual era tratado como una variable algebraica. El planteamiento moderno que vamos a ver aquí (aunque restringido a funciones continuas) data de 1920.

Definición 3.16 Dada una función continua $f :]0, +\infty[\rightarrow \mathbb{R}$, definimos su *transformada de Laplace* como la función⁴

$$\mathcal{L}\{f\}(s) = F(s) = \int_0^{+\infty} f(t)e^{-st} dt.$$

La transformada de Laplace es, pues, una integral impropia (en $+\infty$ y tal vez en 0) que, por consiguiente, puede converger o no. Aquí vamos a adoptar el convenio siguiente:

En lo sucesivo, sólo calcularemos transformadas de Laplace de funciones continuas $f :]0, +\infty[\rightarrow \mathbb{R}$ que sean absolutamente integrables en $]0, 1]$ (o, equivalentemente, en cualquier intervalo $]0, T]$).

Esto implica la convergencia absoluta de las integrales

$$\int_0^T f(t)e^{-st} dt$$

y el único problema de convergencia se da en $+\infty$.

³Si añadimos la hipótesis de que f no toma valores negativos, entonces, si $m \leq n$, se cumple que $0 \leq h_m(y) \leq h_n(y) \leq h(y)$, y esto implica a su vez que las integrales de las funciones $h_n(y)$ convergen en $[0, +\infty[$, y podemos aplicar el teorema de la convergencia monótona 3.9 en lugar de 3.10, y es el único caso que necesitaremos realmente.

⁴Es útil el convenio de representar la transformada de una función con la misma letra, pero mayúscula en vez de minúscula.

Esta condición se cumple, en particular, si $f : [0, +\infty[\rightarrow \mathbb{R}$ es continua, pero podemos admitir que la integral de f sea impropia en 0 con tal de que la integral de $|f(t)|$ también sea convergente.

Observemos en primer lugar que si la transformada converge absolutamente para un cierto s_0 , entonces converge absolutamente para todo $s \geq s_0$, pues

$$|f(t)e^{-st}| = |f(t)e^{-s_0t}| e^{-(s-s_0)t} \leq |f(t)e^{-s_0t}|,$$

y 3.5 nos da la convergencia absoluta de $\mathcal{L}\{f\}(s)$ a partir de la de $\mathcal{L}\{f\}(s_0)$.

Por lo tanto, podemos definir la *abscisa de convergencia absoluta* σ_a de la transformada de Laplace de una función f como el ínfimo de los números reales s en los que ésta converge absolutamente, entendiendo que $\sigma_a = +\infty$ si no converge absolutamente en ningún punto y que $\sigma_a = -\infty$ si converge absolutamente en todos los números reales.

Así una transformada de Laplace converge absolutamente en el intervalo $]\sigma_a, +\infty[$ y tal vez también en σ_a .

Funciones exponencialmente acotadas Una condición suficiente para que la transformada de Laplace de una función continua f tenga abscisa de convergencia absoluta $\sigma_a < +\infty$ es que esté *exponencialmente acotada*, es decir, que existan constantes M y a tales que

$$|f(t)| \leq Me^{at},$$

en cuyo caso $\sigma_a \leq a$, es decir, que la transformada converge absolutamente al menos en el intervalo $]a, +\infty[$.

En efecto, en tal caso $|f(t)e^{-st}| \leq Me^{-(s-a)t}$ y la integral

$$\begin{aligned} \int_0^{+\infty} Me^{-(s-a)t} dt &= \lim_{T \rightarrow +\infty} M \int_0^T e^{-(s-a)t} dt = \lim_{T \rightarrow +\infty} -\frac{M}{s-a} [e^{-(s-a)t}]_0^T \\ &= \lim_{T \rightarrow +\infty} \frac{M}{s-a} (1 - e^{-(s-a)T}) = \frac{M}{s-a} \end{aligned}$$

es convergente si $s > a$. Con esto hemos probado además que, para funciones exponencialmente acotadas,

$$\lim_{s \rightarrow +\infty} F(s) = 0.$$

En particular, para $a = 0$, la transformada de Laplace de una función acotada converge al menos en $]0, +\infty[$. ■

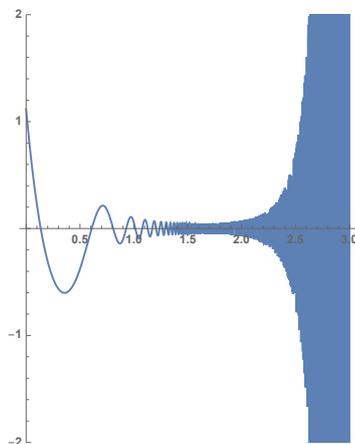
Ahora bien, cabe señalar que una transformada de Laplace puede converger sólo condicionalmente.

Ejemplo *La transformada de Laplace*

$$F(s) = \int_0^{+\infty} e^{e^t} \operatorname{sen} e^{e^t} e^{-st} dt$$

converge condicionalmente para $s > -1$ y no converge absolutamente en ningún punto.

La figura muestra el integrando de $F(5)$:



Aplicamos el cambio de variable

$$u = e^{e^t}, \quad du = e^{e^t} e^t dt = u \log u dt.$$

Así:

$$\int_0^T e^{e^t} \operatorname{sen} e^{e^t} e^{-st} dt = \int_e^{e^{e^T}} \frac{\operatorname{sen} u}{\log^{s+1} u} du,$$

luego

$$\int_0^{+\infty} e^{e^t} \operatorname{sen} e^{e^t} e^{-st} dt = \int_e^{+\infty} \frac{\operatorname{sen} u}{\log^{s+1} u} du,$$

donde la integral de la derecha converge por el criterio de Dirichlet siempre que $s > -1$. Podemos aplicar el mismo cambio de variable a la integral con el seno en valor absoluto, y en el ejemplo de la página 158 hemos visto que la integral no es absolutamente convergente. ■

La utilidad de la transformada de Laplace se debe principalmente a que, por una parte, existen reglas sencillas de transformación que, no sólo permiten calcular fácilmente las transformadas de algunas funciones, sino que en muchos casos permiten resolver el problema inverso de identificar funciones cuya transformada de Laplace sea una función dada. El lector puede ojear la tabla 3.1, que contiene las reglas de este tipo que vamos a demostrar, para hacerse una idea más precisa sobre de qué estamos hablando.

Por otra parte, la transformada de Laplace transforma ciertas manipulaciones de una función f dada en otras manipulaciones de su transformada F , la más importante de las cuales es que, cuando f es derivable y continua en 0, se cumple que

$$\mathcal{L}\{f'\}(s) = s\mathcal{L}\{f\} - f(0),$$

y esto (junto con la propiedad análoga para derivadas sucesivas) permite, como ya habíamos señalado, transformar ecuaciones diferenciales en ecuaciones algebraicas. La tabla 3.2 recoge los resultados de este tipo que vamos a demostrar.

Ejemplo Aceptando las propiedades de la transformada de Laplace que demostraremos más tarde, vamos a resolver la ecuación diferencial

$$y'' + 2y' + y = x, \quad y(0) = 0, \quad y'(0) = 1.$$

para que el lector se forme una idea preliminar de lo que es una “aplicación típica” de la transformada. Si llamamos $Y = \mathcal{L}\{y\}(s)$, tenemos que

$$s^2Y - 1 + 2(sY - 1) + Y = \frac{1}{s^2},$$

donde, con la numeración de la tabla 3.2, hemos aplicado la linealidad (1) para calcular la transformada de cada término por separado, la propiedad (5) sobre transformación de las derivadas para calcular las transformadas de y'' e y' y la entrada (2) de la tabla 3.1 para calcular la transformada del término independiente. Así hemos convertido la ecuación diferencial en una ecuación algebraica en la que podemos despejar Y :

$$(s^2 + 2s + 1)Y = 3 + \frac{1}{s^2},$$

luego

$$Y = \frac{3}{(s+1)^2} + \frac{1}{s^2(s+1)^2} = \frac{2}{s} - \frac{1}{s^2} - \frac{2}{s+1} + \frac{2}{(s+1)^2},$$

donde hemos descompuesto la expresión en fracciones simples porque así es fácil identificar una función cuya transformada de Laplace sea Y . En efecto, sin más que usar la fila (4) de la tabla 3.1 junto con la linealidad, vemos que

$$Y = \mathcal{L}\{y\}(s) = \mathcal{L}\{2 - t - 2e^{-t} + 2te^{-t}\}(s).$$

Por último, según veremos, una función está determinada por su transformada de Laplace, por lo que podemos concluir que

$$y(t) = 2 - t - 2e^{-t} + 2te^{-t}. \quad \blacksquare$$

3.5.1 Cálculo de transformadas de Laplace

El caso (1) de la tabla 3.1 es inmediato:

- Si $f(t) = 1$, entonces $F(s) = 1/s$, para $s > 0$.

En efecto,

$$F(s) = \int_0^{+\infty} e^{-st} dt = \lim_{t \rightarrow +\infty} \int_0^t e^{-st} dt = \lim_{t \rightarrow +\infty} -\frac{1}{s} [e^{-st}]_0^t =$$

$$\lim_{t \rightarrow +\infty} -\frac{1}{s} (e^{-st} - 1) = \frac{1}{s},$$

donde la convergencia sólo se da para $s > 0$.

Una simple inducción nos da el caso (2):

- Si n es un número natural y $f(t) = t^n$, entonces $F(s) = n!/s^{n+1}$, para $s > 0$.

El caso $n = 0$ es el apartado precedente. Supuesto cierto para n , consideramos $f(t) = t^{n+1}$ e integramos por partes:

$$\int t^{n+1} e^{-st} dt = -\frac{t^{n+1}}{s} e^{-st} + \frac{n+1}{s} \int t^n e^{-st} dt,$$

luego

$$\begin{aligned} F(s) &= \int_0^{+\infty} t^{n+1} e^{-st} dt = \lim_{t \rightarrow +\infty} -\frac{t^{n+1}}{s} e^{-st} + \frac{n+1}{s} \int_0^t t^n e^{-st} dt = \\ &= \frac{n+1}{s} \int_0^{+\infty} t^n e^{-st} dt = \frac{n+1}{s} \frac{n!}{s^{n+1}} = \frac{(n+1)!}{s^{n+2}}, \end{aligned}$$

donde el límite del sumando exterior a la integral se ve que es nulo, por ejemplo, mediante la regla de L'Hôpital.

Veamos ahora unos casos particulares de los casos (3) y (4):

- Si $f(t) = \operatorname{sen} t$, entonces $F(s) = 1/(s^2 + 1)$, para $s > 0$.

En efecto,

$$F(s) = \int_0^{+\infty} \operatorname{sen} t e^{-st} dt$$

y basta tener en cuenta (C.5). Igualmente se razona con el coseno:

- Si $f(t) = \operatorname{cos} t$, entonces $F(s) = s/(s^2 + 1)$, para $s > 0$.

Terminamos este apartado con un ejemplo más sofisticado:

- Si $f(t) = 1/\sqrt{t}$, entonces $F(s) = \sqrt{\pi/s}$, para $s > 0$.

Tenemos que estudiar la integral

$$F(s) = \int_0^{+\infty} \frac{1}{\sqrt{t}} e^{-st} dt.$$

Notemos que es absolutamente convergente en $[0, 1]$, pues el integrando está acotado en $]0, 1]$ por la función $1/\sqrt{t}$, que es integrable, mientras que en $[1, +\infty[$ lo es por el criterio de Dirichlet. El cambio de variable $u = st$ transforma la integral en

$$\frac{1}{\sqrt{s}} \int_0^{+\infty} \frac{1}{\sqrt{u}} e^{-u} du,$$

y el cambio $u = x^2$ la transforma a su vez en

$$\frac{2}{\sqrt{s}} \int_0^{+\infty} e^{-x^2} dx,$$

pero esta integral es la integral de Gauss calculada en la página 164. Así pues,

$$F(s) = \sqrt{\frac{\pi}{s}}.$$

Notemos que la linealidad de la transformada de Laplace (la propiedad (1) de la tabla 3.2, que demostraremos justo a continuación) implica que la transformada de $f(t) = 1/\sqrt{\pi t}$ es $F(s) = 1/\sqrt{s}$.

3.5.2 Propiedades de la transformada de Laplace

Veamos ahora algunas de las propiedades de la tabla 3.2, que a su vez implican algunas de las de la tabla 3.1 cuya demostración tenemos pendiente. La propiedad (1) de linealidad es inmediata:

$$\begin{aligned} \mathcal{L}\{af + bg\}(s) &= \int_0^{+\infty} (af(t) + bg(t)) e^{-st} dt = \\ &= a \int_0^{+\infty} f(t) e^{-st} dt + b \int_0^{+\infty} g(t) e^{-st} dt = a \mathcal{L}\{f\}(s) + b \mathcal{L}\{g\}(s). \end{aligned}$$

La propiedad (2) implica el caso general de los casos (3) y (4) de la tabla 3.1 a partir de los casos particulares que hemos demostrado en el apartado anterior:

$$\mathcal{L}\{f(at)\}(s) = \int_0^{+\infty} f(at) e^{-st} dt = \frac{1}{a} \int_0^{+\infty} f(t) e^{-st/a} dt = \frac{1}{a} F(s/a).$$

Hemos hecho el cambio de variable $t' = at$.

Los casos (5), (6), (7) y (8) de la tabla 3.1 se deducen inmediatamente de la propiedad (3) de la tabla 3.2, que probamos a continuación:

$$\mathcal{L}\{e^{at} f(t)\}(s) = \int_0^{+\infty} e^{at} f(t) e^{-st} dt = \int_0^{+\infty} f(t) e^{-(s-a)t} dt = F(s-a).$$

Veamos ya una primera propiedad no trivial, que es la propiedad (4) de la tabla 3.2:

Teorema 3.17 Sea $f :]0, +\infty[\rightarrow \mathbb{R}$ una función continua de modo que

$$\phi(t) = \int_0^t f(\tau) d\tau,$$

sea absolutamente convergente para todo $t > 0$. Si $\mathcal{L}\{f\}(s)$ converge en $s_0 > 0$, entonces $\mathcal{L}\{\phi\}(s)$ también converge en s_0 y

$$\mathcal{L}\{\phi\}(s) = \frac{1}{s} \mathcal{L}\{f\}(s)$$

para todo $s \geq s_0$. Además $\mathcal{L}\{\phi\}(s)$ converge absolutamente para $s > s_0$.

DEMOSTRACIÓN: Definimos

$$\psi(x) = \int_0^x \phi(t)e^{-s_0 t} dt, \quad g(x) = e^{s_0 x} \psi(x), \quad h(x) = e^{s_0 x}.$$

La función $\phi(t)$ es derivable, y en particular continua, luego $\psi(x)$ y $g(x)$ también son derivables (y, por supuesto, $h(x)$ también). La hipótesis $s_0 > 0$ implica que $\lim_{x \rightarrow +\infty} h(x) = +\infty$, luego también $\lim_{x \rightarrow +\infty} g(x) = +\infty$, ya que $\psi(x)$ está acotada.

Vamos a aplicar la regla de L'Hôpital, para lo cual calculamos:

$$\frac{g'(x)}{h'(x)} = \frac{e^{s_0 x}(s_0 \psi(x) + \psi'(x))}{e^{s_0 x} s_0} = \frac{1}{s_0} \left(s_0 \int_0^x \phi(t)e^{-s_0 t} dt + \phi(x)e^{-s_0 x} \right).$$

Integramos por partes:

$$\begin{aligned} u &= \phi(t), \quad du = f(t) dt, \quad dv = e^{-s_0 t} dt, \quad v = -\frac{1}{s_0} e^{-s_0 t} : \\ &= \frac{1}{s_0} \left([-\phi(t)e^{-s_0 t}]_0^x + \int_0^x f(t)e^{-s_0 t} dt + e^{-s_0 x} \phi(x) \right) = \frac{1}{s_0} \int_0^x f(t)e^{-s_0 t} dt. \end{aligned}$$

Como $\mathcal{L}\{f\}(s_0)$ es convergente, la última expresión tiende a $F(s_0)/s_0$, de modo que

$$\mathcal{L}\{\phi\}(s_0) = \lim_{x \rightarrow +\infty} \psi(x) = \lim_{x \rightarrow +\infty} \frac{g(x)}{h(x)} = \lim_{x \rightarrow +\infty} \frac{g'(x)}{h'(x)} = \frac{1}{s_0} F(s_0).$$

Con esto hemos probado el teorema para s_0 , pero todo $s > s_0$ cumple obviamente las mismas hipótesis que s_0 , luego el teorema vale para todo $s \geq s_0$. En particular, hemos demostrado que

$$\lim_{x \rightarrow +\infty} \frac{s_0 \psi(x) + \psi'(x)}{s_0} = \mathcal{L}\{\phi\}(s_0),$$

que equivale a

$$\mathcal{L}\{\phi\}(s_0) + \lim_{x \rightarrow +\infty} \frac{\psi'(x)}{s_0} = \mathcal{L}\{\phi\}(s_0),$$

de donde

$$0 = \lim_{x \rightarrow +\infty} \psi'(x) = \lim_{x \rightarrow +\infty} \phi(x)e^{-s_0 x}.$$

Esto implica que $\phi(x)$ cumple está exponencialmente acotada, por lo que su transformada converge absolutamente para $s > s_0$. ■

De aquí deducimos la propiedad correspondiente a la derivación, que es la pieza clave en la reducción de ecuaciones diferenciales a ecuaciones algebraicas:

Teorema 3.18 Si $f : [0, +\infty[\rightarrow \mathbb{R}$ es continua, derivable en $]0, +\infty[$ con derivada continua y $\mathcal{L}\{f'\}(s)$ converge en un punto $s_0 > 0$, entonces la transformada $\mathcal{L}(f)$ converge en s_0 y

$$\mathcal{L}\{f'\}(s) = s\mathcal{L}\{f\}(s) - f(0)$$

para todo $s \geq s_0$. Además, $\mathcal{L}\{f\}(s)$ converge absolutamente en todo $s > s_0$.

DEMOSTRACIÓN: Basta aplicar el teorema anterior a la función $f'(t)$, de modo que

$$\phi(t) = \int_0^t f'(\tau) d\tau = f(t) - f(0),$$

y tenemos la convergencia para $s \geq s_0$ de

$$\mathcal{L}\{\phi\}(s) = \mathcal{L}\{f\}(s) - f(0)\mathcal{L}\{1\}(s) = \mathcal{L}\{f\}(s) - \frac{f(0)}{s},$$

donde la convergencia de $\mathcal{L}\{\phi\}(s)$ y la de $\mathcal{L}\{1\}(s)$ implican la de $\mathcal{L}\{f\}(s)$, y además se cumple la relación

$$\mathcal{L}\{f\}(s) - \frac{f(0)}{s} = \frac{1}{s}\mathcal{L}\{f'\}(s),$$

que al despejar nos da la relación del enunciado. Teniendo en cuenta

$$|f(t)| \leq |f(t) - f(0)| + |f(0)|,$$

de la convergencia absoluta de $\mathcal{L}\{\phi\}(s)$ se sigue la de $\mathcal{L}\{f\}(s)$. ■

Aplicando sucesivamente este teorema a una función con n derivadas continuas, obtenemos la propiedad (5) de la tabla 3.2:

$$\mathcal{L}\{f^n\}(s) = s^n \mathcal{L}\{f\}(s) - s^{n-1}f(0) - s^{n-2}f'(0) - \dots - sf^{(n-2)}(0) - f^{(n-1)}(0).$$

Veamos ahora la propiedad (6):

Teorema 3.19 Sea $f : [0, +\infty[\rightarrow \mathbb{R}$ una función continua tal que $\mathcal{L}\{f(t)/t\}(s)$ tenga abscisa de convergencia absoluta $\sigma_a < +\infty$. Entonces

$$\mathcal{L}\{f(t)/t\}(s) = \int_s^{+\infty} F(\sigma) d\sigma.$$

DEMOSTRACIÓN: Para todo $s > 0$, el teorema 3.14 nos da la igualdad

$$\int_s^{+\infty} \int_0^{+\infty} f(t)e^{-\sigma t} dt d\sigma = \int_0^{+\infty} \int_s^{+\infty} f(t)e^{-\sigma t} d\sigma dt.$$

Para justificarla tenemos que probar la convergencia de las integrales

$$\int_s^{+\infty} |f(t)|e^{-\sigma t} d\sigma, \quad \int_0^{+\infty} \int_s^{+\infty} |f(t)|e^{-\sigma t} d\sigma dt.$$

La primera es simplemente

$$|f(t)| \int_s^{+\infty} e^{-\sigma t} d\sigma = \frac{|f(t)|}{t} e^{-st}$$

y la segunda es

$$\int_0^{+\infty} \int_s^{+\infty} |f(t)|e^{-\sigma t} d\sigma dt = \int_0^{+\infty} \frac{|f(t)|}{t} e^{-st} dt,$$

que es convergente para $s > \sigma_a$.

El mismo cálculo eliminando el valor absoluto nos da que

$$\mathcal{L}\{f(t)/t\}(s) = \int_0^{+\infty} \int_s^{+\infty} f(t) e^{-\sigma t} d\sigma dt = \int_s^{+\infty} \int_0^{+\infty} f(t) e^{-\sigma t} dt d\sigma = \int_s^{+\infty} F(\sigma) d\sigma.$$

■

Nota Conviene observar que para aplicar el teorema anterior hay que comprobar que la integral

$$\int_0^1 \frac{|f(t)|}{t} e^{-st} dt$$

es convergente, que es una condición necesaria para que esté definida la transformada $\mathcal{L}\{f(t)/t\}$. Por ejemplo, $f(t) = 1$ no cumple este requisito y, correspondientemente, tampoco existe

$$\int_s^{+\infty} F(\sigma) = \int_s^{+\infty} \frac{1}{\sigma} d\sigma.$$

■

Ejemplo: La integral de Dirichlet

$$\int_0^{+\infty} \frac{\operatorname{sen} x}{x} dx = \frac{\pi}{2}.$$

En efecto, si aplicamos el teorema anterior a $f(t) = \operatorname{sen} t$, entonces

$$F(s) = \frac{1}{1+s^2},$$

$$\int_s^{+\infty} F(\sigma) d\sigma = \int_s^{+\infty} \frac{1}{1+\sigma^2} d\sigma = \lim_{S \rightarrow +\infty} [\arctan S]_s^S = \frac{\pi}{2} - \arctan s.$$

El teorema afirma que

$$\int_0^{+\infty} \frac{\operatorname{sen} t}{t} e^{-st} dt = \frac{\pi}{2} - \arctan s.$$

(Notemos que la condición de convergencia absoluta en $[0, 1]$ se cumple trivialmente porque el integrando es continuo.) En muchos libros se lee que ahora basta hacer $s = 0$ en la igualdad anterior para obtener el resultado, pero no es obvio en absoluto que la integral tenga que tender a la integral del enunciado. Eso lo probaremos a partir del teorema siguiente. ■

Teorema 3.20 Si $f : [0, +\infty[\rightarrow \mathbb{R}$ es continua y existe $\lim_{t \rightarrow +\infty} f(t)$, entonces $\lim_{s \rightarrow 0^+} sF(s) = \lim_{t \rightarrow +\infty} f(t)$.

DEMOSTRACIÓN: Ante todo, observemos que f está acotada, por lo que $F(s)$ está definida en $]0, +\infty[$. Llamemos $L = \lim_{t \rightarrow +\infty} f(t)$. Teniendo en cuenta que

$$s \int_0^{+\infty} e^{-st} dt = 1$$

(notemos que la integral es $\mathcal{L}\{1\} = 1/s$), podemos expresar

$$|sF(s) - L| = \left| s \int_0^{+\infty} f(t)e^{-st} dt - s \int_0^{+\infty} L e^{-st} dt \right| \leq s \int_0^{+\infty} |f(t) - L| e^{-st} dt.$$

Tomamos A tal que si $t > A$ entonces $|f(t) - L| < \epsilon/2$, y partimos la integral:

$$|sF(s) - L| \leq 2Ms \int_0^A dt + \frac{\epsilon}{2} s \int_A^{+\infty} e^{-st} dt \leq 2MA s + \frac{\epsilon}{2},$$

luego, dado $\epsilon > 0$, podemos tomar $\delta = \epsilon/4MA$, de modo que si $0 < s < \delta$ entonces $|sF(s) - L| < \epsilon$. ■

Con esto podemos probar el resultado que necesitamos para completar el cálculo de la integral de Dirichlet:

Teorema 3.21 Si $f :]0, +\infty[\rightarrow \mathbb{R}$ es una función continua tal que la integral

$$\int_0^{+\infty} f(t) dt$$

es convergente, entonces

$$\lim_{s \rightarrow 0^+} \int_0^{+\infty} f(t)e^{-st} dt = \int_0^{+\infty} f(t) dt.$$

DEMOSTRACIÓN: La función $g(t) = \int_0^t f(\tau) d\tau$ cumple $g'(t) = f(t)$, luego, integrando por partes:

$$s \int_0^t g(\tau)e^{-s\tau} d\tau = [-g(\tau)e^{-s\tau}]_0^t + \int_0^t f(\tau)e^{-s\tau} d\tau.$$

Por consiguiente,

$$sG(s) = s \int_0^{+\infty} g(t)e^{-st} dt = \int_0^{+\infty} f(t)e^{-st} dt.$$

La función $g(t)$ está en las hipótesis del teorema anterior, luego

$$\lim_{s \rightarrow 0^+} \int_0^{+\infty} f(t)e^{-st} dt = \lim_{s \rightarrow 0^+} sG(s) = \lim_{t \rightarrow +\infty} g(t) = \int_0^{+\infty} f(t) dt. \quad \blacksquare$$

Terminamos esta sección probando un resultado complementario al del teorema 3.20:

Teorema 3.22 Si $f : [0, +\infty[\rightarrow \mathbb{R}$ es continua y exponencialmente acotada, entonces

$$\lim_{s \rightarrow +\infty} sF(s) = f(0).$$

DEMOSTRACIÓN: Supongamos en primer lugar que f está acotada, digamos que $|f(t)| \leq M$ para todo $t \geq 0$. Entonces

$$sF(s) = s \int_0^{+\infty} f(t)e^{-st} dt = \int_0^{+\infty} f(t/s)e^{-t} dt.$$

Teniendo en cuenta que

$$\int_0^{+\infty} e^{-t} dt = 1,$$

para todo $A > 0$, se cumple que

$$\begin{aligned} |sF(s) - f(0)| &= \left| \int_0^{+\infty} f(t/s)e^{-t} dt - \int_0^{+\infty} f(0)e^{-t} dt \right| \leq \\ &\int_0^A |f(t/s) - f(0)| e^{-t} dt + \int_A^{+\infty} |f(t/s) - f(0)| e^{-t} dt \leq \\ &\int_0^A |f(t/s) - f(0)| e^{-t} dt + 2M \int_A^{+\infty} e^{-t} dt. \end{aligned}$$

Dado $\epsilon > 0$, elegimos A tal que

$$\int_A^{+\infty} e^{-t} dt < \frac{\epsilon}{4M}$$

y a su vez elegimos $\delta > 0$ tal que si $|x| < \delta$, entonces $|f(x) - f(0)| < \epsilon/2A$. De este modo, si $s > A/\delta$, se cumple $|t/s| < A/s < \delta$, luego $|f(t/s) - f(0)| < \epsilon/2A$. Así:

$$|sF(s) - f(0)| < \frac{\epsilon}{2A} \int_0^A dt + \frac{\epsilon}{2} = \epsilon.$$

Esto prueba el teorema en el caso en que f está acotada. En el caso general tenemos que $|f(t)| \leq Me^{at}$, y consideramos la función $g(t) = e^{-at}f(t)$, que está acotada, luego podemos aplicar el caso ya probado para concluir que

$$f(0) = g(0) = \lim_{s \rightarrow +\infty} sG(s) = \lim_{s \rightarrow +\infty} sF(s+a),$$

donde hemos usado la propiedad (3) de la tabla 3.2, ya demostrada. Concluimos observando que

$$\lim_{s \rightarrow +\infty} sF(s) = \lim_{s \rightarrow +\infty} (s+a)F(s+a) = f(0) + \lim_{s \rightarrow +\infty} aF(s+a) = f(0),$$

donde hemos usado que, como hemos demostrado en la página 172 si f está exponencialmente acotada, $\lim_{s \rightarrow +\infty} F(s) = 0$. ■

3.5.3 Convergencia condicional

Hasta ahora sólo hemos visto un ejemplo de transformada de Laplace con convergencia condicional. No es un caso frecuente, pero vamos a probar un teorema que permite tratarlo con cierta comodidad:

Teorema 3.23 *Sea $f :]0, +\infty[\rightarrow \mathbb{R}$ una función continua. Si su transformada de Laplace converge en un punto $s = s_0$, entonces converge en todo el intervalo $]s_0, +\infty[$, y además puede expresarse como una integral absolutamente convergente:*

$$F(s) = (s - s_0) \int_0^{+\infty} e^{-(s-s_0)t} \phi(t) dt,$$

donde

$$\phi(t) = \int_0^t e^{-s_0\tau} f(\tau) d\tau.$$

DEMOSTRACIÓN: Por hipótesis existe $\lim_{t \rightarrow +\infty} \phi(t) = F(s_0)$, y $\phi(t)$ es una función continua, de donde se sigue inmediatamente que está acotada en $[0, +\infty[$. Pongamos que $|\phi(t)| \leq M$ para todo $t \geq 0$. Fijado $s > s_0$, integrando por partes:

$$\int_0^\tau e^{-st} f(t) dt = e^{-(s-s_0)\tau} \phi(\tau) + (s - s_0) \int_0^\tau e^{-(s-s_0)t} \phi(t) dt.$$

Además, $|e^{-(s-s_0)\tau} \phi(\tau)| \leq M e^{-(s-s_0)\tau}$, luego

$$\lim_{\tau \rightarrow +\infty} e^{-(s-s_0)\tau} \phi(\tau) = 0$$

y, por el mismo motivo, la última integral converge absolutamente si $\tau \rightarrow +\infty$, lo que nos permite concluir que

$$\int_0^{+\infty} e^{-st} f(t) dt = (s - s_0) \int_0^{+\infty} e^{-(s-s_0)t} \phi(t) dt. \quad \blacksquare$$

Así pues, podemos definir la *abscisa de convergencia* de una transformada de Laplace como el ínfimo $\sigma_c \leq \sigma_a$ de todos los números reales donde es convergente (entendiendo, como en el caso de la abscisa de convergencia absoluta, que es $\sigma_c = -\infty$ si converge en todos los números reales y $\sigma_c = +\infty$ si no converge en ninguno), y entonces el intervalo de convergencia es $] \sigma_c, +\infty[$ (teniendo en cuenta que en σ_c la transformada puede converger o no).

Veamos un ejemplo de la importancia de la representación que proporciona el teorema anterior de las transformadas de Laplace como integrales absolutamente convergentes:

Teorema 3.24 *Sea $f :]0, +\infty[\rightarrow \mathbb{R}$ una función continua cuya transformada de Laplace converja en un punto s_0 . Entonces, para todo $s > s_0$, se cumple que $\mathcal{L}\{tf(t)\}(s) = -F'(s)$.*

DEMOSTRACIÓN: Explícitamente, queremos probar la igualdad

$$F'(s) = - \int_0^{+\infty} t f(t) e^{-st} dt.$$

Esto sería inmediato si pudiéramos aplicar el teorema 3.11 a

$$F(s) = \int_0^{+\infty} f(t) e^{-st} dt,$$

lo cual sucede, por ejemplo, si la función $f(t)$ está acotada. Para el caso general expresamos la transformada según el teorema 3.23:

$$F(s) = (s - s_0) \int_0^{+\infty} \phi(t) e^{-(s-s_0)t} dt, \quad \text{con} \quad \phi(s) = \int_0^t f(\tau) e^{-s_0\tau} d\tau.$$

Derivando resulta

$$F'(s) = \int_0^{+\infty} \phi(t) e^{-(s-s_0)t} dt - (s - s_0) \int_0^{+\infty} t \phi(t) e^{-(s-s_0)t} dt.$$

Ahora bien, para que este paso sea válido tenemos que justificar que el teorema 3.11 es aplicable, para lo cual tomamos $s_0 < s_1 < s$ y observamos que

$$|t \phi(t) e^{-(s-s_0)t}| \leq t M e^{-(s_1-s_0)t} = \phi^*(t),$$

donde M es una cota de $|\phi(t)|$ y claramente

$$\int_0^{+\infty} \phi^*(t) dt = M \mathcal{L}\{t\}(s_1 - s_0) = \frac{1}{(s_1 - s_0)^2}.$$

Por lo tanto, la derivación está justificada, y vamos a probar que la expresión que hemos obtenido coincide con la del enunciado. Para ello integramos por partes:

$$\int_0^t \tau e^{-s_0\tau} f(\tau) d\tau = [\tau \phi(\tau)]_0^t - \int_0^t \phi(\tau) d\tau,$$

luego

$$t \phi(t) = \int_0^t (\phi(\tau) + \tau e^{-s_0\tau} f(\tau)) d\tau.$$

De nuevo integramos por partes:

$$\begin{aligned} & -(s - s_0) \int_0^{+\infty} t \phi(t) e^{-(s-s_0)t} dt = \\ & \lim_{u \rightarrow +\infty} \left([t \phi(t) e^{-(s-s_0)t}]_0^u - \int_0^u e^{-(s-s_0)t} (\phi(t) + t f(t) e^{-s_0t}) dt \right) \\ & = - \int_0^{+\infty} e^{-(s-s_0)t} \phi(t) dt - \int_0^{+\infty} t f(t) e^{-st} dt, \end{aligned}$$

y al sustituir esto en la expresión que hemos obtenido para $F'(s)$ queda la expresión del enunciado. ■

Aplicando repetidamente este teorema obtenemos la propiedad (7) de la tabla 3.2.

Con los casos (6), (7) y (8) ya demostrados de la tabla 3.1 es claro que siempre podemos encontrar una función f cuya transformada sea una fracción simple cualquiera cuyo denominador no tenga una raíz compleja múltiple. El teorema anterior nos permite abordar este caso:

Ejemplo Vamos a encontrar una función $f(t)$ cuya transformada de Laplace sea

$$F(s) = \frac{1}{(s^2 + 1)^2}.$$

Para ello observamos que

$$F(s) = -\frac{1}{2s} \frac{d}{ds} \frac{1}{s^2 + 1},$$

luego si llamamos $g(t) = \sin t$, sabemos que $G(s) = 1/(s^2 + 1)$, luego 3.24 nos da que $\mathcal{L}\{-t \sin t\}(s) = G'(s)$ y 3.17 nos da que $F(s)$ es la transformada de Laplace de

$$f(t) = \frac{1}{2} \int_0^t \tau \sin \tau \, d\tau = \frac{1}{2}(\sin t - t \cos t). \quad \blacksquare$$

Uno de los resultados más importantes que tenemos pendiente demostrar es que funciones distintas tienen transformadas de Laplace distintas. Para ello necesitamos un resultado previo:

Teorema 3.25 Si $f : [a, b] \rightarrow \mathbb{R}$ es una función continua y

$$\int_a^b x^n f(x) \, dx = 0$$

para todo número natural n , entonces f es idénticamente nula.

DEMOSTRACIÓN: Dado $\epsilon > 0$, por el teorema de aproximación de Weierstrass 2.4, podemos tomar un polinomio $p(x)$ tal que $|f(x) - p(x)| < \epsilon$, para todo $x \in [a, b]$. Equivalentemente, $f(x) = p(x) + \epsilon\theta(x)$, donde $\theta : [a, b] \rightarrow [-1, 1]$ es una función continua. Multiplicando por $f(x)$ e integrando obtenemos que

$$\int_a^b f(x)^2 \, dx = \int_a^b f(x)p(x) \, dx + \epsilon \int_a^b f(x)\theta(x) \, dx.$$

Ahora bien, si $p(x) = \sum_{n=0}^r c_n x^n$, por hipótesis tenemos que

$$\int_a^b f(x)p(x) \, dx = \sum_{n=0}^r c_n \int_a^b x^n f(x) \, dx = 0,$$

luego

$$\int_a^b f(x)^2 \, dx = \left| \int_a^b f(x)^2 \, dx \right| = \left| \epsilon \int_a^b f(x)\theta(x) \, dx \right| \leq \epsilon \int_a^b |f(x)| \, dx.$$

Si f no fuera idénticamente nula, por continuidad sería estrictamente positiva en un intervalo, luego la primera y la última integral de la línea anterior serían estrictamente positivas, y podríamos escribir

$$0 < \frac{\int_a^b f(x)^2 dx}{\int_a^b |f(x)| dx} \leq \epsilon,$$

lo cual es absurdo, porque ϵ puede tomarse arbitrariamente pequeño. ■

Ahora podemos probar un resultado muy general de unicidad:

Teorema 3.26 Si $f : [0, +\infty[\rightarrow \mathbb{R}$ es una función continua cuya transformada de Laplace converge en un punto s_0 y se anula en todos los puntos de la forma $s_0 + ns_1$, con $s_1 > 0$, entonces f es idénticamente nula.

DEMOSTRACIÓN: Según 3.23, para todo $s > s_0$, tenemos que

$$F(s) = (s - s_0) \int_0^{+\infty} e^{-(s-s_0)t} \phi(t) dt,$$

y la hipótesis es que, para todo número natural n ,

$$F(s_0 + ns_1) = ns_1 \int_0^{+\infty} e^{-ns_1 t} \phi(t) dt = 0.$$

Consideramos el cambio de variable

$$x = e^{-s_1 t}, \quad t = -\frac{\log x}{s_1}, \quad dt = -\frac{1}{s_1 x} dx,$$

con lo que

$$0 = \int_0^1 x^{n-1} \phi(-s_1^{-1} \log x) dx.$$

Notemos que la función $\phi(-s_1^{-1} \log x)$ es continua en $[0, 1]$, pues en 0 puede ser definida como

$$\lim_{x \rightarrow 0} \phi(-s_1^{-1} \log x) = \lim_{t \rightarrow +\infty} \phi(t) = \int_0^{+\infty} e^{-s_0 \tau} f(\tau) d\tau = F(s_0).$$

El teorema 3.25 implica que la aplicación $\phi(-s_1^{-1} \log x)$ es idénticamente nula en el intervalo $[0, 1]$, lo cual equivale a que $\phi(t)$ es idénticamente nula en el intervalo $[0, +\infty[$. Así pues, para todo $t > 0$ se cumple que

$$\int_0^t e^{-s_0 \tau} f(\tau) d\tau = 0.$$

Derivando respecto de t , resulta que $e^{-s_0 t} f(t) = 0$, luego $f(t) = 0$, para todo $t > 0$. ■

En particular:

Teorema 3.27 Si dos funciones continuas $f, g : [0, +\infty[\rightarrow \mathbb{R}$ tienen la misma transformada de Laplace (o, más en general, si sus transformadas de Laplace coinciden sobre una sucesión $s_0 + ns_1$, con $s_1 > 0$), entonces son iguales.

Basta tener en cuenta que la transformada de Laplace de $f - g$ es la diferencia de las transformadas de Laplace, luego se anula en la sucesión indicada, y el teorema anterior nos da que $f - g = 0$.

Con esto ya tenemos completamente justificada la resolución de la ecuación diferencial de la página 174. Veamos otro más:

Ejemplo Vamos a resolver el sistema de ecuaciones diferenciales

$$x' = y + e^t, \quad y' = -x + \operatorname{sen} t, \quad x(0) = 1, \quad y(0) = 0.$$

El teorema 3.18 y la tabla 3.1 nos permiten transformar fácilmente el sistema en un sistema de ecuaciones algebraicas:

$$sX - 1 = Y + \frac{1}{s-1}, \quad sY = -X + \frac{s}{s^2+1}.$$

Equivalentemente:

$$sX - Y = \frac{s}{s-1}, \quad X + sY = \frac{1}{s^2+1}.$$

Multiplicando una ecuación por s y sumándole o restándole la otra queda:

$$(s^2+1)X = \frac{s^2}{s-1} + \frac{1}{s^2+1}, \quad (s^2+1)Y = \frac{s}{s^2+1} - \frac{s}{s-1}.$$

Despejando y descomponiendo en fracciones simples queda

$$\begin{aligned} X &= \frac{1}{2} \left(\frac{1}{s-1} + \frac{1}{s^2+1} + \frac{s}{s^2+1} \right) + \frac{1}{(s^2+1)^2}, \\ Y &= \frac{s}{(s^2+1)^2} - \frac{1}{2} \left(\frac{1}{s-1} + \frac{1}{s^2+1} - \frac{s}{s^2+1} \right). \end{aligned}$$

Y ahora la tabla 3.1, el ejemplo de la página 184 y el teorema 3.24 nos dan inmediatamente las funciones cuyas transformadas de Laplace son estas dos funciones:

$$x(t) = \frac{1}{2}(e^t + \operatorname{sen} t + \cos t) + \frac{1}{2}(\operatorname{sen} t - t \cos t) = \frac{1}{2}(e^t + 2 \operatorname{sen} t + \cos t - t \cos t),$$

$$y(t) = \frac{1}{2}t \operatorname{sen} t - \frac{1}{2}(e^t + \operatorname{sen} t - \cos t) = -\frac{1}{2}(e^t + \operatorname{sen} t - \cos t - t \operatorname{sen} t).$$

Notemos que $s/(s^2+1)^2$ es la derivada de $-\mathcal{L}\{\operatorname{sen} t\}(s)$. ■

3.5.4 Convolución

Veamos una última propiedad de las que hacen particularmente útil a la transformada de Laplace. Necesitamos una definición:

Definición 3.28 Si $f, g : [0, +\infty[\rightarrow \mathbb{R}$ son dos funciones continuas, definimos su *convolución* como la función $f * g : [0, +\infty[\rightarrow \mathbb{R}$ dada por

$$(f * g)(x) = \int_0^x f(x-t)g(t) dt.$$

Observemos que el cambio de variable $t' = x - t$ prueba que $f * g = g * f$.

Se cumple que $f * g$ es una función continua, pues, dado $x_0 \geq 0$, tomamos una cota M de f y g en el intervalo $[0, x_0 + 1]$ y, por el teorema de Heine-Cantor [ITAn .14], dado $\epsilon > 0$, podemos tomar $\delta > 0$ tal que si $x, y \in [0, x_0 + 1]$ cumplen $|x - y| < \delta$, entonces $|f(x) - f(y)| < \epsilon/2Mx_0$. Podemos suponer además que $\delta < \epsilon/2M^2$. Entonces, si $|x_0 - x| < \delta$,

$$\begin{aligned} |(f * g)(x) - (f * g)(x_0)| &= \left| \int_{x_0}^x f(x-t)g(t) dt + \int_0^{x_0} (f(x-t) - f(x_0-t))g(t) dt \right| \\ &\leq \pm \int_{x_0}^x |f(x-t)||g(t)| dt + \int_0^{x_0} |f(x-t) - f(x_0-t)||g(t)| dt, \end{aligned}$$

donde el doble signo será negativo si $x < x_0$. Ahora observamos que

$$|(x-t) - (x_0-t)| = |x - x_0| < \delta,$$

luego

$$|(f * g)(x) - (f * g)(x_0)| \leq M^2|x - x_0| + \frac{\epsilon}{2Mx_0}Mx_0 < \epsilon,$$

luego $f * g$ es continua en x_0 . ■

El resultado fundamental sobre convoluciones es el teorema siguiente, que se corresponde con la propiedad (8) de la tabla 3.2:

Teorema 3.29 $f, g : [0, +\infty[\rightarrow \mathbb{R}$ son dos funciones continuas cuyas transformadas de Laplace convergen absolutamente para $s > s_0$, entonces

$$\mathcal{L}\{f * g\}(s) = \mathcal{L}\{f\}(s)\mathcal{L}\{g\}(s).$$

DEMOSTRACIÓN: Si llamamos F y G a las transformadas de Laplace de f y g , respectivamente, tenemos que

$$\begin{aligned} F(s)G(s) &= \int_0^{+\infty} e^{-st}f(t) dt \int_0^{+\infty} e^{-su}g(u) du = \\ &= \int_0^{+\infty} \int_0^{+\infty} e^{-s(t+u)}f(t)g(u) dt du. \end{aligned}$$

En la integral interior hacemos el cambio de variable $x = t + u$, $dx = dt$, con lo que

$$F(s)G(s) = \int_0^{+\infty} \int_u^{+\infty} e^{-sx} f(x-u)g(u) dx du.$$

Pero todos los cálculos que hemos hecho valen también cambiando f y g por $|f|$ y $|g|$, lo que nos da la convergencia de las integrales

$$\int_u^{+\infty} |e^{-sx} f(x-u)g(u)| dx, \quad \int_0^{+\infty} \int_u^{+\infty} |e^{-sx} f(x-u)g(u)| dx du.$$

Por lo tanto, podemos aplicar el teorema 3.15 para concluir⁵ que

$$F(s)G(s) = \int_0^{+\infty} \int_0^x e^{-sx} f(x-u)g(u) du dx.$$

Equivalentemente:

$$F(s)G(s) = \int_0^{+\infty} e^{-sx} \int_0^x f(x-u)g(u) du dx = \mathcal{L}(f * g)\{s\}. \quad \blacksquare$$

Nota El teorema anterior, junto con el hecho de que cada función está determinada por su transformada de Laplace, implica trivialmente que la convolución cumple propiedades análogas al producto de funciones:

$$f * g = g * f, \quad f * (g * h) = (f * g) * h, \quad f * (g + h) = f * g + f * h. \quad \blacksquare$$

3.6 La función factorial

En la introducción hemos contado que Euler demostró que la integral paramétrica

$$\int_0^1 (-\log t)^s dt$$

interpola a la sucesión $n!$. Gauss aplicó el cambio de variable $x = -\log t$, con lo que $t = e^{-x}$ y $dt = -e^{-x} dx$ y en 1813 definió la *función factorial*

$$\Pi(s) = \int_0^{+\infty} x^s e^{-x} dx.$$

Teorema 3.30 *La integral*

$$\Pi(s) = \int_0^{+\infty} x^s e^{-x} dx$$

es convergente para todo número real $s > -1$ y cumple la ecuación funcional

$$\Pi(s+1) = (s+1)\Pi(s).$$

En particular, $\Pi(n) = n!$ para todo número natural n .

⁵Según se señala en la nota al pie en la prueba de dicho teorema, si f y g no toman valores negativos la prueba puede basarse en el teorema de la convergencia monótona en lugar de en el teorema de la convergencia dominada.

Tabla 3.1: Transformadas de Laplace de algunas funciones elementales

	$f(t)$	$F(s)$	pág.
1	1	$\frac{1}{s}$	174
2	t^n	$\frac{n!}{s^{n+1}}$	175
3	$\text{sen } \omega t$	$\frac{\omega}{s^2 + \omega^2}$	175
4	$\text{cos } \omega t$	$\frac{s}{s^2 + \omega^2}$	175
5	e^{-at}	$\frac{1}{s+a}$	176
6	$t^n e^{-at}$	$\frac{n!}{(s+a)^{n+1}}$	176
7	$e^{-at} \text{sen } \omega t$	$\frac{\omega}{(s+a)^2 + \omega^2}$	176
8	$e^{-at} \text{cos } \omega t$	$\frac{s}{(s+a)^2 + \omega^2}$	176
8	t^α	$\frac{\Gamma(\alpha)}{s^{\alpha+1}}$	190
9	$\log t$	$-\frac{\log s + \gamma}{s}$	196

Tabla 3.2: Propiedades de la transformada de Laplace

	$f(t)$	$F(s)$	pág.
1	$af + bg$	$aF + bG$	176
2	$f(at)$	$\frac{1}{a}F(s/a)$	176
3	$e^{at}f(t)$	$F(s-a)$	176
4	$\int_0^t f(\tau) d\tau$	$F(s)/s$	176
5	$f^{(n)}(t)$	$s^n F - \sum_{i=0}^{n-1} s^i f^{(n-1-i)}(0)$	178
6	$f(t)/t$	$\int_0^s F(\sigma) d\sigma$	178
7	$t^n f(t)$	$-F^{(n)}(s)$	183
8	$(f * g)(t)$	$F(s)G(s)$	187

$$\lim_{s \rightarrow 0^+} sF(s) = \lim_{t \rightarrow +\infty} f(t), \quad \lim_{s \rightarrow +\infty} sF(s) = f(0).$$

DEMOSTRACIÓN: En primer lugar veremos que la integral converge en el intervalo $[1, +\infty[$. Como el integrando es una función continua en $]0, +\infty[$, sabemos que tiene primitiva en $[1, +\infty[$, luego la integral sólo es impropia en $+\infty$. Observemos que

$$\lim_{x \rightarrow +\infty} \frac{x^s e^{-x}}{1/x^2} = \lim_{x \rightarrow +\infty} \frac{x^{s+2}}{e^x} = 0,$$

por ejemplo acotando $s + 2 \leq n$ por un número natural y aplicando n veces la regla de L'Hôpital. Por lo tanto, existe un $M > 0$ tal que si $x > M$ entonces $x^s e^{-x} \leq 1/x^2$ y, como la integral de esta función converge en $[M, +\infty[$, el teorema 3.5 nos da que la de $x^s e^{-x}$ también lo hace, y claramente lo mismo vale para $[1, +\infty[$.

En el intervalo $[0, 1]$, si $s \geq 0$ el integrando es continuo en \mathbb{R} , luego la integral no es impropia, mientras que si $-1 < s < 0$, entonces $0 \leq x^s e^{-x} \leq x^s$ y la función x^s determina una integral convergente por 3.2, luego la integral de $x^s e^{-x}$ también es convergente por el teorema 3.5.

Para probar la ecuación funcional aplicamos la fórmula de integración por partes:

$$\int x^{s+1} e^{-x} dx = -x^{s+1} e^{-x} + (s+1) \int x^s e^{-x} dx.$$

Si la aplicamos al intervalo $[1, t]$ queda

$$\int_1^t x^{s+1} e^{-x} dx = e^{-1} - t^{s+1} e^{-t} + (s+1) \int_1^t x^s e^{-x} dx,$$

de donde

$$\int_1^{+\infty} x^{s+1} e^{-x} dx = e^{-1} + (s+1) \int_1^{+\infty} x^s e^{-x} dx,$$

donde hemos usado de nuevo que $\lim_{t \rightarrow +\infty} t^{s+1}/e^t = 0$. Similarmente, en el intervalo $[t, 1]$ tenemos que

$$\int_t^1 x^{s+1} e^{-x} dx = t^{s+1} e^{-t} - e^{-1} + (s+1) \int_t^1 x^s e^{-x} dx,$$

de donde

$$\int_0^1 x^{s+1} e^{-x} dx = -e^{-1} + (s+1) \int_0^1 x^s e^{-x} dx,$$

y sumando las dos integrales queda la ecuación funcional.

Es fácil ver que $\Pi(0) = 1$, con lo que un razonamiento inductivo a partir de la ecuación funcional implica inmediatamente que $\Pi(n) = n!$. ■

Ahora podemos generalizar la línea (2) de la tabla 3.1 de transformadas de Laplace y demostrar la línea (8) (aplicamos el cambio de variable $st = x$):

$$\mathcal{L}\{t^\alpha\}(s) = \int_0^{+\infty} t^\alpha e^{-st} dt = \int_0^{+\infty} x^\alpha s^{-(\alpha+1)} e^{-x} dx = \frac{\Pi(\alpha)}{s^{\alpha+1}}.$$

En particular $\Pi(\alpha) = \mathcal{L}\{t^\alpha\}(1)$, y el cálculo de la transformada de Laplace de $1/\sqrt{t}$ que hemos hecho en la página 175 se traduce ahora en que

$$\Pi(-1/2) = \sqrt{\pi}, \quad \Pi(1/2) = \frac{\sqrt{\pi}}{2}.$$

La segunda igualdad sale de sustituir $s = -1/2$ en la ecuación funcional.

La figura muestra la gráfica de la función factorial. Observamos que

$$\lim_{x \rightarrow -1} \Pi(s) = +\infty,$$

lo cual es consecuencia inmediata de la ecuación funcional, si la ponemos en la forma

$$\Pi(s) = \frac{\Pi(s+1)}{s+1}.$$

Notemos que para calcular el límite en -1 de esta expresión necesitamos usar que la función $\Pi(s)$ es continua (en 0). Vamos a probar que, de hecho, es una función derivable en $] -1, +\infty[$. Más en general, vamos a usar el teorema 3.11 para probar que sus derivadas sucesivas vienen dadas por

$$\Pi^{(n)}(s) = \int_0^{+\infty} x^s e^{-x} \log^n x \, dx. \quad (3.4)$$

Vamos a aplicarlo por separado en los intervalos $[0, 1]$ y $[1, +\infty[$. No perdemos generalidad si restringimos s a un intervalo $-1 < s_0 < s < s_1$. Por una parte,

$$\int_1^{+\infty} x^s e^{-x} \log^n x \, dx \leq \int_1^{+\infty} x^{s_1+n} e^{-x} \, dx,$$

donde usamos que $\log x \leq x$, luego podemos aplicar el teorema 3.11 tomando $\phi(x) = x^{s_1+n} e^{-x}$, cuya integral converge porque es un tramo de la integral que define a $\Pi(s_1+n)$, que ya hemos probado que es convergente. Por otra parte,

$$\int_0^1 x^s e^{-x} |\log^n x| \, dx \leq \int_0^1 x^{s_0} (-\log x)^n \, dx,$$

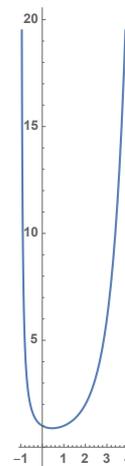
y en este caso basta tomar $\phi(x) = x^{s_0} (-\log x)^n$, cuya integral converge porque el cambio de variable $u = -\log x$ la transforma en

$$\int_0^1 u^n e^{-(s_0+1)u} \, du,$$

y a su vez el cambio $(s_0+1)u = x$ la transforma en

$$\frac{1}{(s_0+1)^{n+1}} \int_0^{+\infty} x^n e^{-x} \, dx = \frac{\Pi(n)}{(s_0+1)^{n+1}}.$$

Vamos a probar ahora dos curiosas fórmulas descubiertas en 1697 por Johann Bernoulli:



Teorema 3.31

$$\int_0^1 \frac{1}{x^x} dx = \sum_{n=1}^{\infty} \frac{1}{n^n} = \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \frac{1}{4^4} + \dots$$

$$\int_0^1 x^x dx = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^n} = \frac{1}{1^1} - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^4} + \dots$$

DEMOSTRACIÓN: Tenemos que $x^{-x} = e^{-x \log x}$. La serie de potencias de la exponencial nos da

$$\frac{1}{x^x} = \sum_{n=0}^{\infty} \frac{(-1)^n x^n \log^n x}{n!}.$$

Ahora observamos que, para $0 \leq x \leq 1$ la sucesión

$$\sum_{n=0}^N \frac{(-1)^n x^n \log^n x}{n!}$$

tiene todos sus sumandos positivos, luego es creciente y converge a x^{-x} . Podemos aplicar el teorema de la convergencia monótona 3.9 para concluir que

$$\begin{aligned} \int_0^1 \frac{1}{x^x} dx &= \int_0^1 \lim_N \sum_{n=0}^N \frac{(-1)^n x^n \log^n x}{n!} = \lim_N \int_0^1 \sum_{n=0}^N \frac{(-1)^n x^n \log^n x}{n!} \\ &= \lim_N \sum_{n=0}^N \int_0^1 \frac{(-1)^n x^n \log^n x}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^1 x^n \log^n x dx. \end{aligned}$$

Para calcular las integrales aplicamos el cambio de variable $[0, +\infty[\rightarrow]0, 1]$ dado por $x = e^{-t/(n+1)}$, cuyo cambio inverso es $t = -(n+1) \log x$. Fijado $0 < x_0 < 1$, se cumple:

$$\begin{aligned} \int_{x_0}^1 x^n \log^n x dx &= \int_{-(n+1) \log x_0}^0 e^{-nt/(n+1)} \frac{(-1)^n t^n}{(n+1)^n} e^{-t/(n+1)} \frac{-1}{n+1} dt \\ &= \frac{(-1)^{n+1}}{(n+1)^{n+1}} \int_{-(n+1) \log x_0}^0 t^n e^{-t} dt = \frac{(-1)^n}{(n+1)^{n+1}} \int_0^{-(n+1) \log x_0} t^n e^{-t} dt. \end{aligned}$$

y tomando el límite cuando $x_0 \rightarrow 0$ queda

$$\int_0^1 x^n \log^n x dx = \frac{(-1)^n}{(n+1)^{n+1}} \int_0^{+\infty} t^n e^{-t} dt = \frac{(-1)^n}{(n+1)^{n+1}} \Gamma(n+1) = \frac{(-1)^n n!}{(n+1)^{n+1}}.$$

Por lo tanto,

$$\int_0^1 \frac{1}{x^x} dx = \sum_{n=0}^{\infty} \frac{1}{(n+1)^{n+1}} = \sum_{n=1}^{\infty} \frac{1}{n^n}.$$

La segunda fórmula se demuestra de forma idéntica sin más que suprimir el $(-1)^n$ del desarrollo en serie de partida. La única modificación es que ahora

no podemos usar el teorema de la convergencia monótona para intercambiar la serie y la integral porque la sucesión de sumas parciales no es creciente. En su lugar podemos usar el teorema de la convergencia dominada, pero vamos a dar un argumento alternativo que vale en ambos casos.

Para ello observamos que la función $x \log x$ es continua en $[0, 1]$, por lo que existe un M tal que $|x \log x| \leq M$ para todo x en dicho intervalo. (Aunque es irrelevante, es fácil ver que la función toma su valor mínimo en $x = 1/e$, donde vale $-1/e$, luego sirve $M = 1/e$.) Entonces

$$\begin{aligned} & \left| \sum_{n=0}^N \int_0^1 \frac{x^n \log^n x}{n!} dx - \int_0^1 \sum_{n=0}^{\infty} \frac{x^n \log^n x}{n!} dx \right| = \\ & \left| \int_0^1 \left(\sum_{n=0}^N \frac{x^n \log^n x}{n!} - \sum_{n=0}^{\infty} \frac{x^n \log^n x}{n!} \right) dx \right| \leq \\ & \int_0^1 \left| \sum_{n=0}^N \frac{x^n \log^n x}{n!} - \sum_{n=0}^{\infty} \frac{x^n \log^n x}{n!} \right| dx = \int_0^1 \left| \sum_{n=N+1}^{\infty} \frac{x^n \log^n x}{n!} \right| dx \leq \\ & \int_0^1 \sum_{n=N+1}^{\infty} \frac{|x \log x|^n}{n!} dx \leq \int_0^1 \sum_{n=N+1}^{\infty} \frac{M^n}{n!} dx = \sum_{n=N+1}^{\infty} \frac{M^n}{n!}, \end{aligned}$$

y la última sucesión tiende a 0 con n , luego

$$\int_0^1 \sum_{n=0}^{\infty} \frac{x^n \log^n x}{n!} dx = \lim_N \sum_{n=0}^N \int_0^1 \frac{x^n \log^n x}{n!} dx = \sum_{n=0}^{\infty} \int_0^1 \frac{x^n \log^n x}{n!} dx.$$

■

Ahora vamos a probar que la función factorial que hemos definido es la misma que introdujimos en [ITAn 8.16]. Recordemos que allí la definimos como

$$\Pi(z) = \lim_n \frac{n! n^z}{(z+1) \cdots (z+n)},$$

para todo número complejo z que no sea un entero negativo. Salvo un factor $(n+1)^z/n^z$, que tiende a 1 con n , esta expresión es la primera que obtuvo Euler para interpolar la sucesión de factoriales (véase la introducción).

Además, en [ITAn] demostramos que la función factorial allí definida cumple la ecuación funcional $\Pi(z+1) = (z+1)\Pi(z)$. Esto nos da una expresión alternativa ligeramente distinta:

$$\Pi(z) = \frac{\Pi(z+1)}{z+1} = \lim_n \frac{n! n^{z+1}}{(z+1) \cdots (z+n+1)}.$$

Lo que vamos a demostrar es que si $s > -1$ es un número real, entonces la función $\Pi(s)$ que hemos definido aquí cumple

$$\Pi(s) = \lim_n \frac{n! n^{s+1}}{(s+1) \cdots (s+n+1)},$$

por lo que ciertamente es la restricción al intervalo $]-1, +\infty[$ de la función definida en [ITAn 8.16].

Para ello recordemos [ITAn 4.6] que

$$e^{-x} = \lim_n \left(1 - \frac{x}{n}\right)^n.$$

Veamos en primer lugar que

$$\Pi(s) = \int_0^{+\infty} \lim_n x^s \left(1 - \frac{x}{n}\right)^n dx = \lim_n \int_0^n x^s \left(1 - \frac{x}{n}\right)^n dx.$$

Para ello haremos algunas manipulaciones que nos sitúen en las hipótesis del teorema 2.8. Dado $\epsilon > 0$, podemos tomar $M, \delta > 0$ tales que

$$\int_1^{+\infty} x^s e^{-x} dx - \int_1^M x^s e^{-x} dx < \frac{\epsilon}{4}, \quad \int_0^1 x^s e^{-x} dx - \int_\delta^1 x^s e^{-x} dx < \frac{\epsilon}{4}.$$

Si probamos que existe un $n_0 \geq M$ tal que si $n \geq n_0$ se cumple

$$\int_\delta^M x^s e^{-x} dx - \int_\delta^M x^s \left(1 - \frac{x}{n}\right)^n dx < \frac{\epsilon}{2},$$

podremos concluir que

$$\begin{aligned} \Pi(s) - \int_0^n x^s e^{-x} dx &= \int_0^\delta x^s e^{-x} dx + \int_\delta^M x^s e^{-x} dx + \int_M^{+\infty} x^s e^{-x} dx \\ &\quad - \int_0^\delta x^s \left(1 - \frac{x}{n}\right)^n dx - \int_\delta^M x^s \left(1 - \frac{x}{n}\right)^n dx - \int_M^n x^s \left(1 - \frac{x}{n}\right)^n dx \\ &\leq \frac{\epsilon}{4} + \frac{\epsilon}{4} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

y el teorema estará demostrado. Equivalentemente, tenemos que probar que

$$\int_\delta^M x^s e^{-x} dx = \lim_n \int_\delta^M x^s \left(1 - \frac{x}{n}\right)^n dx.$$

Para ello basta probar que las funciones $f_n : [\delta, M] \rightarrow \mathbb{R}$ dadas por

$$f_n(x) = x^s \left(1 - \frac{x}{n}\right)^n$$

y $f(x) = x^s e^{-x}$ cumplen las condiciones del teorema 3.9. Podemos considerar únicamente índices $n > M$, de modo que si $x \in [\delta, M]$ se cumple $0 < x/n < 1$. Ciertamente tenemos que $\lim_n f_n(x) = f(x)$. Además, por [ITAn 4.5], se cumple que

$$1 - \frac{x}{n} < e^{-x/n},$$

luego

$$\left(1 - \frac{x}{n}\right)^n < e^{-x}$$

y $f_n(x) < x^s e^{-x}$.

Falta probar que cada sucesión $\{f_n(x)\}_n$ es creciente. En efecto:

$$\begin{aligned} \frac{f_{n+1}(x)}{f_n(x)} &= \left(\frac{n+1-x}{n+1}\right)^{n+1} \left(\frac{n}{n-x}\right)^n = \left(\frac{n+1-x}{n+1} \frac{n}{n-x}\right)^{n+1} \frac{n-x}{n} \\ &= \left(\frac{n^2+n-nx}{n^2+n-nx-x}\right)^{n+1} \frac{n-x}{n} = \left(1 + \frac{x}{(n+1)(n-x)}\right)^{n+1} \frac{n-x}{n} \\ &> \left(1 + \frac{x}{n-x}\right) \frac{n-x}{n} = \frac{n-x}{n} + \frac{x}{n} = 1, \end{aligned}$$

donde en la desigualdad hemos sustituido la potencia del binomio por los dos primeros términos del desarrollo de Newton, aprovechando que todos los términos suprimidos son positivos.

Así pues,

$$\Pi(s) = \lim_n \frac{1}{n^n} \int_0^n (1-x)^n x^s dx.$$

Vamos a aplicar la fórmula de integración por partes. Hay que tener en cuenta que $s > -1$, por lo que la integral es impropia en 0 cuando $-1 < s < 0$. Para $t > 0$ tenemos que

$$\int_t^n (n-x)^n x^s dx = -\frac{(n-t)^{n+1} t^{s+1}}{s+1} + \frac{n}{s+1} \int_t^n (n-x)^{n-1} x^{s+1} dx$$

y al tomar el límite cuando $t \rightarrow 0$ queda que

$$\int_0^n (n-x)^n x^s dx = \frac{n}{s+1} \int_0^n (n-x)^{n-1} x^{s+1} dx,$$

donde la integral de la derecha ya no es impropia en ningún caso. Aplicando $n-1$ veces más la fórmula de integración por partes llegamos a que

$$\begin{aligned} \Pi(s) &= \lim_n \frac{1}{n^n} \frac{n!}{(s+1)(s+2)\cdots(s+n)} \int_0^n x^{s+n} dx \\ &= \lim_n \frac{1}{n^n} \frac{n!}{(s+1)(s+2)\cdots(s+n)} \frac{n^{s+n+1}}{s+n+1} = \lim_n \frac{n! n^{s+1}}{(s+1)\cdots(s+n+1)}, \end{aligned}$$

como queríamos probar. ■

Así pues, según [ITAn 8.16], tenemos también una expresión para la función factorial en términos de un producto infinito:

$$\Pi(x) = e^{-\gamma x} \prod_{k=1}^{\infty} \left(1 + \frac{x}{k}\right)^{-1} e^{x/k},$$

donde γ es la constante de Euler-Mascheroni. El teorema [ITAn 8.3] nos da que

$$\log \Pi(x) = -\gamma x + \sum_{k=1}^{\infty} \left(\frac{x}{k} - \log\left(1 + \frac{x}{k}\right)\right),$$

y podemos aplicar el teorema 1.21 para derivar la serie:

$$\frac{\Pi'(x)}{\Pi(x)} = -\gamma + \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{k+x} \right) = -\gamma + \sum_{k=1}^{\infty} \frac{x}{k(k+x)}.$$

En efecto, para ello observamos que, si $-1 < -1 + \delta < x < r$, con $\delta > 0$ y $r \geq 1$, entonces, para $k \geq 2$,

$$\frac{|x|}{k(k+x)} \leq \frac{r}{k(k-1)} = M_k$$

mientras que, para $k = 1$, es

$$\frac{|x|}{x+1} \leq \max\left\{\frac{1}{\delta}, r\right\} = M_1$$

y la serie

$$\sum_{k=1}^{\infty} M_k = M_1 + r \sum_{k=2}^{\infty} \frac{1}{k(k-1)} = M_1 + r$$

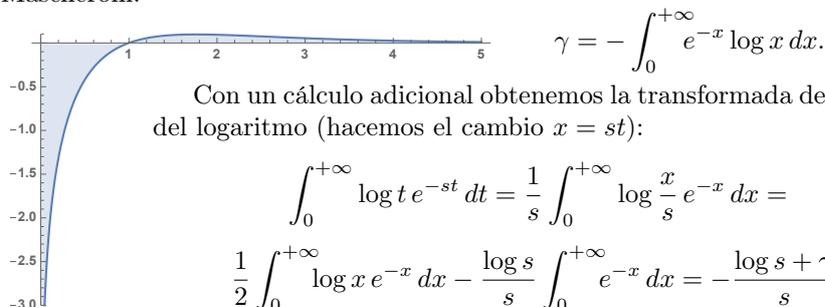
es convergente (por [ITAn], ejemplo 2.5). Esto prueba que la expresión para la derivada del logaritmo es válida para todo x que cumpla $-1 + \delta < x < r$, luego para todo $x > -1$. Así pues:

$$\Pi'(x) = \Pi(x) \left(-\gamma + \sum_{k=1}^{\infty} \frac{x}{k(k+x)} \right).$$

Si en particular evaluamos en $x = 0$ y $x = 1$ obtenemos que

$$\Pi'(0) = -\gamma, \quad \Pi'(1) = 1 - \gamma$$

y, comparando con la expresión integral para la derivada que nos proporciona la igualdad (3.4), obtenemos una expresión integral para la constante de Euler-Mascheroni:



Las integrales de Euler Lagrange llamó *integral de Euler de segunda especie* a la integral que define la función factorial, mientras que la *integral de Euler de primera especie* era

$$E(x, y) = \int_0^1 t^x (1-t)^y \, dt,$$

que obviamente está definida cuando $x, y \geq 0$, pero que, de hecho, es convergente para $x, y > -1$.

En efecto, sabemos que, para $x, y > -1$, las transformadas de Laplace de las funciones t^x y t^y son, respectivamente, $\Pi(x)/s^{x+1}$, $\Pi(y)/s^{y+1}$, luego el teorema 3.29 nos da que la transformada de Laplace de $t^x * t^y$ es $\Pi(x)\Pi(y)/s^{x+y+2}$. Ahora bien, esta función es claramente la transformada de Laplace de

$$\frac{\Pi(x)\Pi(y)}{\Pi(x+y+1)} t^{x+y+1},$$

luego concluimos que

$$\int_0^u (u-t)^x t^y dt = \frac{\Pi(x)\Pi(y)}{\Pi(x+y+1)} t^{x+y+1}.$$

Ahora basta hacer $u = 1$ para concluir:

Teorema 3.32 *Para todo $x, y > -1$, se cumple que*

$$E(x, y) = \int_0^1 (1-t)^x t^y dt = \frac{\Pi(x)\Pi(y)}{\Pi(x+y+1)}.$$

Esta expresión muestra en particular que $E(x, y) = E(y, x)$.

Para números naturales tenemos que

$$E(m, n) = \frac{m!n!}{(m+n+1)!}.$$

En particular,

$$\binom{n}{m} = \frac{n+1}{E(n-m, m)}.$$

La función $E(x, y)$ permite obtener información sobre la función factorial. Por ejemplo, tomando $x = y = -1/2$ en la fórmula del teorema 3.32 obtenemos un cálculo alternativo de $\Pi(-1/2)$ (hacemos el cambio $t = \sin^2 u$):

$$\Pi(-1/2)^2 = \int_0^1 \frac{dt}{\sqrt{(1-t)t}} = \int_0^{\pi/2} 2 du = \pi,$$

con lo que $\Pi(-1/2) = \sqrt{\pi}$.

Veamos otro ejemplo:

Teorema 3.33 (Fórmula de duplicación de Legendre) *Para $x > -1/2$:*

$$\sqrt{\pi} \Pi(2x) = 2^{2x} \Pi(x-1/2)\Pi(x).$$

DEMOSTRACIÓN: Para cualquier $x > -1$, haciendo el cambio $t = (1+u)/2$, se cumple que

$$\begin{aligned} \frac{\Pi^2(x)}{\Pi(2x+1)} &= E(x, x) = \int_0^1 (1-t)^x t^x dt = \frac{1}{2} \int_{-1}^1 \left(\frac{1-u}{2}\right)^x \left(\frac{1+u}{2}\right)^x du \\ &= \frac{1}{2^{2x+1}} \int_{-1}^1 (1-u^2)^x du = \frac{1}{2^x} \int_0^1 (1-u^2)^x du, \end{aligned}$$

donde en el último paso hemos usado que el cambio $u' = -u$ muestra que la integral en $[-1, 0]$ coincide con la integral en $[0, 1]$.

Por otro lado, el cambio $t = u^2$ nos da que

$$\frac{\Pi(x)\Pi(-1/2)}{\Pi(x+1/2)} = E(x, -1/2) = \int_0^1 (1-t)^x t^{-1/2} dt = 2 \int_0^1 (1-u^2)^x du.$$

Combinando las dos igualdades resulta

$$2^{2x} \Pi^2(x)\Pi(x+1/2) = \Pi(2x+1) \frac{1}{2} \Pi(x)\Pi(-1/2)$$

o equivalentemente

$$2^{2x+1} \Pi(x+1/2)\Pi(x) = (2x+1)\Pi(2x)\sqrt{\pi}$$

Si $x > -1/2$ podemos sustituir $\Pi(x+1/2) = \frac{1}{2}(2x+1)\Pi(x-1/2)$ y llegamos a la expresión del enunciado. ■

Por ejemplo, para $x = 3/4$ la relación de Legendre se reduce a

$$\sqrt{\pi} \Pi(3/2) = \sqrt{8} \Pi(3/4)\Pi(1/4),$$

y usando que $\Pi(3/2) = (3/2)\Pi(1/2) = (3/4)\Pi(-1/2) = 3\sqrt{\pi}/4$, llegamos a que

$$\Pi(1/4)\Pi(3/4) = \frac{3\sqrt{2}\pi}{16}. \quad (3.5)$$

Ejemplo Las integrales de Euler permiten reducir varias integrales a la función factorial. Por ejemplo, el cambio de variable $t = x^4$ nos da:

$$\begin{aligned} \int_0^1 \frac{1}{\sqrt{1-x^4}} dx &= \frac{1}{4} \int_0^1 (1-t)^{-1/2} t^{-3/4} dt = \frac{1}{4} E(-1/2, -3/4) \\ &= \frac{\Pi(-1/2)\Pi(-3/4)}{4\Pi(-1/4)} = \frac{3\sqrt{\pi}\Pi(1/4)}{4\Pi(3/4)} = \frac{4\Pi^2(1/4)}{\sqrt{2\pi}}, \end{aligned} \quad (3.6)$$

donde en el último paso hemos usado (3.5). Similarmente,

$$\begin{aligned} \int_0^1 \frac{x^2}{\sqrt{1-x^4}} dx &= \frac{1}{4} \int_0^1 (1-t)^{-1/2} t^{-1/4} dt = \frac{1}{4} E(-1/2, -1/4) \\ &= \frac{\Pi(-1/2)\Pi(-1/4)}{4\Pi(1/4)} = \frac{\sqrt{\pi}\Pi(3/4)}{3\Pi(1/4)} = \frac{\pi\sqrt{2\pi}}{16\Pi^2(1/4)}, \end{aligned}$$

donde nuevamente hemos usado (3.5). En particular tenemos la relación que Euler obtuvo sobre la elástica rectangular (véase la introducción):

$$\int_0^1 \frac{1}{\sqrt{1-x^4}} dx \int_0^1 \frac{x^2}{\sqrt{1-x^4}} dx = \frac{\pi}{4}. \quad (3.7)$$

■

Las funciones beta y gamma En 1808, Legendre definió la *función gamma* como

$$\Gamma(x) = \Pi(x-1) = \int_0^{+\infty} t^{x-1} e^{-t} dt$$

y representó por (x/y) la integral $E(x, y)$, pero en 1839 Jacques Binet definió la *función beta* como

$$B(x, y) = E(x-1, y-1) = \int_0^1 (1-t)^{x-1} t^{y-1} dt.$$

Con esta notación, la relación del teorema 3.32 se expresa en la forma

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Capítulo IV

La dinámica newtoniana

Los elementos de cálculo diferencial e integral que hemos discutido en los capítulos anteriores son suficientes para exponer los principios básicos de la dinámica Newtoniana. En la primera sección nos ocupamos de la *cinemática*, es decir, de la mera descripción (matemática) del movimiento, mientras que en las secciones siguientes abordaremos la *dinámica* propiamente dicha, es decir, el estudio de las causas del movimiento, que permite predecir cómo se moverán unos objetos en unas condiciones dadas.

4.1 Cinemática

Vamos a considerar un objeto puntual, es decir, del que hacemos abstracción de su forma concreta y consideramos que su posición puede ser descrita simplemente por un punto. Por ejemplo, podemos decir que un coche está en el kilómetro 230 de una carretera, sin que importe qué punto concreto del coche está en esa posición, o que una pelota está a 5 m de altura, sin que importe qué punto de la pelota está concretamente a esa altura, y también podemos considerar que la Tierra es un punto cuando estudiamos su movimiento alrededor del Sol, etc.

Fijado un sistema de referencia (un origen de coordenadas y unos ejes cartesianos), la posición del objeto en cada instante t viene dada por un *vector de posición*¹ $\vec{r}(t) = (x(t), y(t), z(t))$, donde introducimos el convenio de representar con una flecha las magnitudes vectoriales y sin flecha las magnitudes escalares.

Vamos a suponer que la función $\vec{r} : I \rightarrow \mathbb{R}^3$ es derivable, lo cual significa que está definida la *velocidad*

$$\vec{v} = \frac{d\vec{r}}{dt} : I \rightarrow \mathbb{R}^3.$$

¹Cuando el movimiento se produzca en un plano podemos elegir el sistema de referencia de modo que éste sea el de ecuación $z = 0$ y suprimir la última coordenada. Igualmente, en una recta basta usar una coordenada.

En el supuesto de que sea no nula, hemos visto que es siempre un vector tangente a la trayectoria del objeto. Más aún, supondremos que \vec{v} es derivable, con lo que está definida la *aceleración*

$$\vec{a} = \frac{d\vec{v}}{dt} : I \longrightarrow \mathbb{R}^3,$$

que determina la variación de la velocidad por unidad de tiempo. En particular, esto implica que \vec{v} es continua, luego, fijando un instante $t_0 \in I$, podemos definir

$$s(t) = \int_{t_0}^t \|\vec{v}(u)\| du,$$

que, según la definición 2.13 y el teorema y la discusión posterior, se interpreta como el espacio recorrido por el móvil desde el instante t_0 hasta el instante t medido sobre su trayectoria (no como la distancia $\|\vec{r}(t) - \vec{r}(t_0)\|$, medida en línea recta), y con el convenio de que es negativo si $t < t_0$.

Así, no hay que confundir la velocidad \vec{v} con la *velocidad sobre la trayectoria*

$$v = \|\vec{v}\| = \frac{ds}{dt},$$

que indica el espacio recorrido sobre la trayectoria por unidad de tiempo.

Igualmente, no hay que confundir la aceleración \vec{a} con la *aceleración sobre la trayectoria*

$$a = \frac{dv}{dt},$$

que mide la variación de la velocidad v sobre la trayectoria y que, como pronto veremos, no se cumple necesariamente $a = \|\vec{a}\|$.

Movimiento uniforme y uniformemente acelerado Se llama *movimiento (rectilíneo) uniforme* al movimiento con velocidad \vec{v} constante. En tal caso tenemos que²

$$\vec{r}(t) - \vec{r}(t_0) = \int_{t_0}^t \vec{v} du,$$

luego, llamando \vec{r}_0 a la posición en el instante t_0 , tenemos que

$$\vec{r} = \vec{r}_0 + \vec{v}(t - t_0),$$

que es la *ecuación del movimiento rectilíneo uniforme*, que nos permite saber la posición del objeto en cualquier instante conocida su posición \vec{r}_0 en un instante t_0 y su velocidad \vec{v} . Notemos que esta fórmula ya la obtuvimos en la página 29, pero allí suponíamos a priori que el movimiento era rectilíneo, mientras que ahora esto lo obtenemos como consecuencia de que la velocidad es constante.

²En estos cálculos tomamos como integral de una función vectorial el vector formado por las integrales de sus funciones coordenadas, pues así obtenemos ciertamente una función cuyas funciones coordenadas son primitivas de las funciones coordenadas de la función dada.

Un movimiento es *uniformemente acelerado* si la aceleración es constante. Si hacemos con el vector constante \vec{a} el mismo cálculo que acabamos de hacer con \vec{v} resulta la ecuación

$$\vec{v} = \vec{v}_0 + \vec{a}(t - t_0),$$

donde \vec{v}_0 es la velocidad en el instante t_0 . Integrando de nuevo:

$$\vec{r} - \vec{r}_0 = \int_{t_0}^t (\vec{v}_0 + \vec{a}(u - t_0)) du,$$

de donde

$$\vec{r} = \vec{r}_0 + \vec{v}_0(t - t_0) + \frac{1}{2}\vec{a}(t - t_0)^2.$$

Ésta es la *ecuación del movimiento uniformemente acelerado*. El movimiento será rectilíneo (es decir, el que estudiamos en la página 29) si los vectores \vec{v}_0 y \vec{a} tienen la misma dirección. ■

Ejemplo: Un satélite artificial Consideremos un sistema de referencia que tenga su origen en el centro de la Tierra y de modo que el plano que contiene los ejes XY sea el plano ecuatorial. Supongamos que un satélite artificial gira alrededor de la Tierra en este plano, y que su posición en cada instante viene dada por

$$\vec{r}(t) = \left(r \cos \frac{2\pi t}{T}, r \sin \frac{2\pi t}{T} \right).$$

Entonces su velocidad es

$$\vec{v}(t) = \left(-\frac{2\pi r}{T} \sin \frac{2\pi t}{T}, \frac{2\pi r}{T} \cos \frac{2\pi t}{T} \right),$$

y el módulo de la velocidad es en todo momento $v = 2\pi r/T$. Esto significa que el satélite se mueve en una órbita circular de radio r a una “velocidad constante” de $2\pi r/T$ m/s, y que T es el tiempo que tarda en recorrer la longitud $2\pi r$ de una órbita completa, es decir T es el tiempo que tarda en dar una vuelta completa alrededor de la Tierra.

Hemos escrito “velocidad constante” entre comillas por que la velocidad del satélite no es constante (es la función $\vec{v}(t)$). Lo que es constante es la velocidad sobre la trayectoria, y eso significa que el satélite recorre siempre $2\pi r/T$ metros cada segundo, pero no lo hace en línea recta. Si la velocidad $\vec{v}(t)$ fuera constante, es decir, si el vector velocidad apuntara siempre en la misma dirección, el satélite recorrería una recta y no una circunferencia.

El hecho de que la velocidad no sea constante se traduce en que la aceleración no es nula:

$$\vec{a}(t) = \left(-\frac{(2\pi)^2 r}{T^2} \cos \frac{2\pi t}{T}, -\frac{(2\pi)^2 r}{T^2} \sin \frac{2\pi t}{T} \right) = -\left(\frac{2\pi}{T} \right)^2 \vec{r}(t). \quad \blacksquare$$

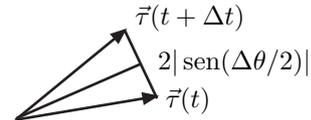
Para distinguir las variaciones del módulo de la velocidad y su dirección, si $\vec{v}(t) \neq \vec{0}$ (equivalentemente, si $v(t) \neq 0$), conviene definir el *vector tangente unitario* como el vector $\vec{\tau}: I \rightarrow \mathbb{R}^3$ dado por

$$\vec{\tau} = \frac{\vec{v}}{v}.$$

Así pues, la velocidad $\vec{v} = v\vec{\tau}$ está determinada por el escalar v que indica la tasa de variación del espacio recorrido en cada instante y el vector $\vec{\tau}$ que indica la dirección de avance en cada instante. Ahora vamos a probar que la aceleración \vec{a} , que indica la tasa de variación de \vec{v} , puede descomponerse en dos partes, una que indica la variación de v y otra que indica la variación de $\vec{\tau}$. La primera es la aceleración sobre la trayectoria $a = dv/dt$ que ya hemos definido.

Por otra parte, la variación de $\vec{\tau}$ está determinada por la derivada $\vec{\tau}'$. Vamos a analizarla con detalle. Fijado un instante t y un incremento pequeño Δt , llamamos $\Delta\theta$ al ángulo que forman los vectores $\vec{\tau}(t + \Delta t)$ y $\vec{\tau}$, de modo que $\Delta\theta$ es una función de Δt . Puesto que $\vec{\tau}$ tiene norma constante igual a 1, la trigonometría nos da que

$$\|\vec{\tau}(t + \Delta t) - \vec{\tau}(t)\| = 2 \left| \sin \frac{\Delta\theta}{2} \right|.$$



Por consiguiente,

$$\left\| \frac{\vec{\tau}(t + \Delta t) - \vec{\tau}(t)}{\Delta t} \right\| = \frac{\|\vec{\tau}(t + \Delta t) - \vec{\tau}(t)\|}{|\Delta t|} = \left| \frac{\sin \frac{\Delta\theta}{2}}{\Delta\theta/2} \right| \frac{\Delta\theta}{|\Delta t|}.$$

Es claro que $\Delta\theta \rightarrow 0$ cuando $\Delta t \rightarrow 0$, luego (usando la continuidad de la norma $\|(x, y, z)\| = \sqrt{x^2 + y^2 + z^2}$):

$$\|\vec{\tau}'(t)\| = \lim_{\Delta t \rightarrow 0} \frac{\Delta\theta}{|\Delta t|}.$$

(Notemos que el hecho de que $\vec{\tau}$ sea derivable implica la existencia del límite de la derecha.) Esto significa que $\|\vec{\tau}'(t_0)\|$ es la tasa de variación del ángulo que $\vec{\tau}(t)$ forma con $\vec{\tau}(t_0)$ por unidad de tiempo.

En este punto conviene hacer un retoque para obtener una magnitud que sólo dependa de la forma de la trayectoria y no de la velocidad a la que se recorra:

Definimos la *curvatura* como

$$\kappa(t) = \frac{\|\vec{\tau}'(t)\|}{v(t)},$$

de modo que $\|\vec{\tau}'\| = v\kappa$.

La clave es que, con esta definición, la curvatura de una curva no se altera si se reparametriza.

En efecto, si tenemos una curva $\vec{r}(t)$ y le aplicamos un cambio de parámetro $t(u)$ para obtener una reparametrización $\vec{r}(u) = \vec{r}(t(u))$, la relación entre las velocidades es

$$\vec{v}(u) = \vec{v}(t(u)) \frac{dt}{du},$$

y tomando módulos

$$v(u) = v(t(u)) \left| \frac{dt}{du} \right|.$$

Los vectores tangentes son

$$\vec{\tau}(u) = \frac{\vec{v}(u)}{v(u)} = \frac{\vec{v}(t(u))}{v(t(u))} (\pm 1) = \pm \vec{\tau}(t(u)),$$

donde el signo es positivo si la reparametrización recorre la curva en el mismo sentido y negativo si la recorre en sentido opuesto. Derivando:

$$\vec{\tau}'(u) = \pm \vec{\tau}'(t(u)) \frac{dt}{du},$$

luego

$$\|\vec{\tau}'(u)\| = \|\vec{\tau}'(t(u))\| \left| \frac{dt}{du} \right|,$$

luego

$$\kappa(u) = \frac{\|\vec{\tau}'(u)\|}{v(u)} = \frac{\|\vec{\tau}'(t(u))\|}{v(t(u))} = \kappa(t(u)).$$

Esto significa que la curvatura de $\vec{r}(t)$ en el punto correspondiente al parámetro $t(u)$ es la misma que la curvatura de $\vec{r}(u)$ en el (mismo) punto correspondiente al parámetro u .

La interpretación geométrica es clara: al dividir la variación del ángulo por unidad de tiempo entre el espacio recorrido $v = ds/st$ por unidad de tiempo, resulta que la curvatura $\kappa(t_0)$ se interpreta como la variación del ángulo que el vector $\vec{\tau}(t)$ forma con $\vec{\tau}(t_0)$ por unidad de arco recorrida.

Más precisamente, suponiendo que la velocidad $\vec{r}'(t)$ no es nunca nula, es decir, que \vec{r} es una curva paramétrica regular, sabemos que podemos construir una reparametrización $\vec{r}(s)$ que esté parametrizada por el arco, es decir, tal que $v = 1$, y entonces, según hemos probado, $\kappa(s) = \|\tau'(t)\|$ es la variación del ángulo de $\vec{\tau}$ por cada unidad que aumenta s , es decir, por cada unidad de arco recorrida.

Por lo tanto, una curvatura grande indica que un pequeño avance sobre el arco se corresponde con una variación grande del vector $\vec{\tau}$, y esto es lo que expresamos cuando decimos que una curva “está muy curvada”.

Ejemplo: La circunferencia Consideremos una circunferencia de radio r , que podemos parametrizar como

$$\vec{r}(t) = (r \cos t, r \sin t).$$

Entonces $\vec{v}(t) = (-r \sin t, r \cos t)$, luego $v = r$ y $\vec{r}'(t) = (-\sin t, \cos t)$. A su vez

$$\vec{r}''(t) = (-\cos t, -\sin t), \quad \|\vec{r}''(t)\| = 1,$$

luego la curvatura es $\kappa = 1/r$.

Esto era previsible: si pensamos en $\vec{r}(t)$ como la trayectoria de un móvil, cada vez que $\vec{r}(t)$ gira 1 radián, el móvil avanza r unidades, luego por cada unidad que avanza el móvil, el vector $\vec{r}(t)$ gira $1/r$ radianes y el vector $\vec{r}''(t)$, que es ortogonal a $\vec{r}(t)$, también gira $1/r$ radianes.

En cambio, la igualdad $\|\vec{r}''(t)\| = 1$ sólo expresa que \vec{r}'' gira un radian por cada unidad que aumenta t , pero el aumento de t no dice nada sobre la circunferencia, ya que depende de la parametrización elegida. En cambio, la curvatura $1/r$ es el giro del vector tangente por cada unidad que nos desplazamos sobre la circunferencia, lo cual es algo intrínseco a ella, independiente de cómo la parametricemos.

Vemos que la curvatura de una circunferencia es menor cuanto mayor es el radio y, en efecto, una circunferencia está menos curvada cuanto mayor es su radio. ■

Ejemplo: La espiral logarítmica Vamos a calcular la curvatura de la espiral logarítmica $\rho = ae^{k\theta}$. Tenemos que $\vec{r} = \rho(\cos \theta, \sin \theta)$, luego

$$\vec{v} = \rho'(\cos \theta, \sin \theta) + \rho(-\sin \theta, \cos \theta) = \rho(k(\cos \theta, \sin \theta) + (-\sin \theta, \cos \theta)),$$

$$v = \rho\sqrt{1+k^2}, \quad \vec{r}' = \frac{1}{\sqrt{1+k^2}}(k(\cos \theta, \sin \theta) + (-\sin \theta, \cos \theta)),$$

$$\vec{r}'' = \frac{1}{\sqrt{1+k^2}}(k(-\sin \theta, \cos \theta) + (-\cos \theta, -\sin \theta)),$$

luego $\|\vec{r}''\| = 1$, luego la definición de curvatura nos da que

$$\kappa = \frac{1}{\rho\sqrt{1+k^2}}.$$

En la página 105 hemos visto que la longitud de la espiral logarítmica viene dada por

$$s = \frac{\sqrt{1+k^2}}{k}\rho,$$

y así vemos que $\kappa s = 1/k$, de modo que, en una espiral logarítmica, la curvatura es inversamente proporcional a la longitud del arco. ■

Hasta aquí tenemos una interpretación del módulo $\|\vec{\tau}'\| = v\kappa$, y nos falta interpretar su dirección. Para ello observamos que, en general, si $\vec{u}(t)$ y $\vec{v}(t)$ son dos funciones vectoriales derivables, la derivada de su producto escalar puede calcularse con la regla del producto:

$$\frac{d(\vec{u} \cdot \vec{v})}{dt} = \frac{d\vec{u}}{dt} \cdot \vec{v} + \vec{u} \cdot \frac{d\vec{v}}{dt}.$$

En efecto, si $\vec{u} = (x, y, z)$ y $\vec{v} = (a, b, c)$, entonces $\vec{u} \cdot \vec{v} = xa + yb + zc$, luego

$$\begin{aligned} \frac{d(\vec{u} \cdot \vec{v})}{dt} &= x'a + xa' + y'b + yb' + z'c + zc' \\ &= (x', y', z')(a, b, c) + (x, y, z)(a', b', c') = \frac{d\vec{u}}{dt} \cdot \vec{v} + \vec{u} \cdot \frac{d\vec{v}}{dt}. \end{aligned}$$

Aplicamos esto al vector $\vec{\tau}$, que cumple $\vec{\tau} \cdot \vec{\tau} = 1$, luego al derivar queda

$$\vec{\tau}' \cdot \vec{\tau} + \vec{\tau} \cdot \vec{\tau}' = 2\vec{\tau}\vec{\tau}' = 0,$$

luego vemos que $\vec{\tau}'$, si es no nulo, es un vector ortogonal a $\vec{\tau}$. En el espacio tridimensional hay infinitas direcciones perpendiculares a una dada. Enseguida veremos cuál de ellas es la que determina $\vec{\tau}'$, pero antes, siempre en el caso en que $\vec{\tau}' \neq \vec{0}$, definimos el *vector normal unitario* como

$$\vec{n} = \frac{\vec{\tau}'}{\|\vec{\tau}'\|} = \frac{1}{v\kappa}\vec{\tau}'.$$

Hemos probado que \vec{n} es ortogonal a $\vec{\tau}$, y $\vec{\tau}' = v\kappa\vec{n}$.

En resumen, tenemos que la aceleración sobre la trayectoria a contiene la información sobre la variación del módulo v de la velocidad y el vector $v\kappa\vec{n}$ contiene la información sobre cómo varía la dirección $\vec{\tau}$ de la velocidad. Ahora mostramos cómo se relacionan estas magnitudes con la aceleración \vec{a} . Para ello derivamos la relación $\vec{v} = v\vec{\tau}$:

$$\vec{a} = a\vec{\tau} + v\vec{\tau}',$$

luego, si $\vec{\tau}' \neq \vec{0}$ llegamos a una igualdad fundamental:

$$\vec{a} = a\vec{\tau} + v^2\kappa\vec{n}$$

Llamaremos *aceleración tangencial* y *aceleración normal* de $\vec{r}(t)$ a los vectores $\vec{a}_T = a\vec{\tau}$ y $\vec{a}_N = \kappa v^2\vec{n}$, respectivamente, de modo que $\vec{a} = \vec{a}_T + \vec{a}_N$.

De este modo hemos descompuesto la aceleración \vec{a} en suma de dos vectores ortogonales, la aceleración tangencial es tangente a la trayectoria, y su módulo es $|a|$. Más precisamente, si \vec{a}_T apunta en el sentido del movimiento es que $a > 0$, luego la velocidad v está aumentando, mientras que si \vec{a}_T apunta en el sentido opuesto al movimiento es que $a < 0$ y la velocidad v está disminuyendo.

Por otro lado, el módulo de la aceleración normal es κv^2 , luego eliminando el factor v^2 obtenemos la curvatura. Así, podemos decir que la aceleración tangencial determina la variación del módulo de la velocidad y la aceleración normal determina la variación de la dirección de la velocidad.

Más aún, ahora podemos precisar que el vector normal unitario \vec{n} es el vector unitario ortogonal a \vec{r} contenido en el mismo plano que el vector \vec{a} y que apunta hacia el semiplano respecto de la recta tangente de dirección \vec{r} hacia el que apunta \vec{a} , que es el semiplano hacia el que se curva la trayectoria.

Observemos que $\vec{n}(t)$ no está definido si y sólo si $\vec{r}'(t) = 0$, lo cual equivale a que $\vec{a}(t) = a(t)\vec{r}(t)$, es decir, a que $\vec{a}(t) = \vec{a}_T(t)$, a que no haya aceleración normal.

Ejemplo: Una pelota de tenis Supongamos que lanzamos una pelota de tenis a 15 m/s con una inclinación de 53° . Esto significa que su velocidad inicial (digamos en $t_0 = 0$) es³

$$\vec{v}_0 = 15(\cos 53^\circ, \operatorname{sen} 53^\circ) \approx (9, 12) \text{ m/s}.$$

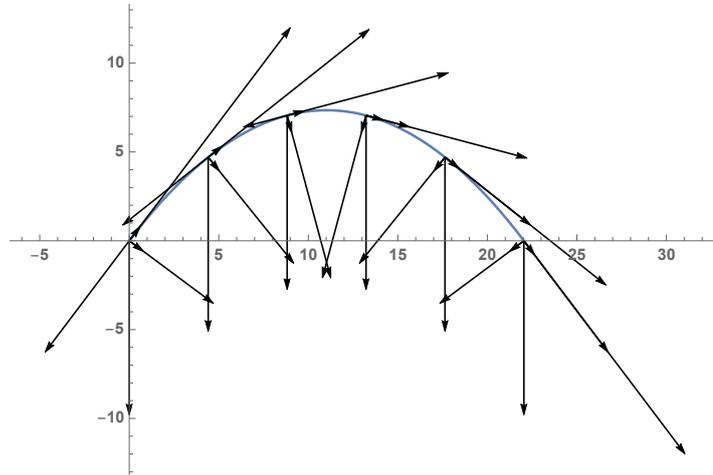
Por otro lado, ya hemos señalado que la aceleración que produce la gravedad es $(0, -9.8)$, por lo que tenemos un movimiento uniformemente acelerado que cumple

$$\vec{r} = (9t, 12t - 4.9t^2), \quad \vec{v} = (9, 12 - 9.8t).$$

La pelota tocará el suelo cuando $12t - 4.9t^2 = 0$. Descartando la solución $t = 0$, que corresponde al momento del saque, obtenemos $t = 12/4.9 = 2.45$ s. El punto donde tocará el suelo estará a $9 \cdot 2.45 = 22.05$ m del punto de partida. Ahora bien, el espacio recorrido por la pelota (sobre su trayectoria) será

$$\int_0^{2.45} \sqrt{225 - 235.2t + 96.04t^2} dt = 27.45 \text{ m}.$$

La integral puede calcularse mediante un cambio de variable debido a Euler, pero los cálculos son bastante laboriosos, y es más práctico dejar que los haga un ordenador.



³Todos los vectores que vamos a considerar serán de la forma $(x, 0, z)$, por lo que omitimos la segunda coordenada.

La figura muestra los vectores \vec{v} y \vec{a} en varios puntos de la trayectoria, así como los vectores unitarios $\vec{\tau}$ y \vec{n} y también \vec{a}_T y \vec{a}_N . Vemos que en el instante inicial la aceleración tangencial es grande (y se opone al movimiento) y la normal es pequeña, lo cual se traduce en que la velocidad sobre la trayectoria desciende rápidamente y la curvatura es pequeña. En cambio, en los puntos más altos es al revés: la aceleración tangencial es pequeña, por lo que la velocidad sobre la trayectoria apenas varía, y la aceleración normal es grande, por lo que la curvatura es grande. ■

Veamos ahora que podemos calcular a , \vec{n} y κ a partir de v , \vec{v} y \vec{a} sin necesidad de calcular más derivadas. En efecto, por el teorema de Pitágoras tenemos que

$$\|\vec{a}\|^2 = a^2 + \kappa^2 v^4,$$

luego la curvatura viene dada por

$$\kappa = \frac{\sqrt{\|\vec{a}\|^2 - a^2}}{v^2}.$$

Más aún,

$$a = v' = (\sqrt{\vec{v} \cdot \vec{v}})' = \frac{2\vec{a} \cdot \vec{v}}{2\sqrt{\vec{v} \cdot \vec{v}}} = \frac{\vec{v} \cdot \vec{a}}{v},$$

luego

$$\kappa^2 = \frac{\|\vec{a}\|^2 - (\vec{v} \cdot \vec{a})^2/v^2}{v^4} = \frac{\|\vec{a}\|^2 v^2 - (\vec{v} \cdot \vec{a})^2}{v^6} = \frac{\|\vec{v} \times \vec{a}\|^2}{v^6},$$

donde hemos usado que si \vec{a} y \vec{v} forman un ángulo α , entonces

$$\|\vec{a}\|^2 v^2 - (\vec{v} \cdot \vec{a})^2 = v^2 \|\vec{a}\|^2 - v^2 \|\vec{a}\|^2 \cos^2 \alpha = v^2 \|\vec{a}\|^2 \sin^2 \alpha = \|\vec{v} \times \vec{a}\|^2.$$

Por lo tanto:

$$\kappa = \frac{\|\vec{v} \times \vec{a}\|}{v^3}.$$

Finalmente, la aceleración tangencial es

$$\vec{a}_T = a\vec{\tau} = \frac{\vec{v} \cdot \vec{a}}{v} \frac{\vec{v}}{v} = \frac{(\vec{v} \cdot \vec{a})\vec{v}}{v^2},$$

luego la aceleración normal es

$$\vec{a}_N = \vec{a} - \frac{(\vec{v} \cdot \vec{a})\vec{v}}{v^2} = \kappa v^2 \vec{n} = \frac{\|\vec{v} \times \vec{a}\|}{v} \vec{n},$$

luego

$$\vec{n} = \frac{v^2 \vec{a} - (\vec{v} \cdot \vec{a})\vec{v}}{v \|\vec{v} \times \vec{a}\|}.$$

En resumen:

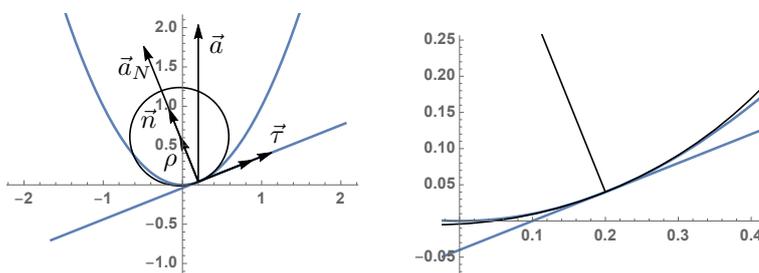
$$\vec{\tau} = \frac{\vec{v}}{v}, \quad \vec{n} = \frac{v^2 \vec{a} - (\vec{v} \cdot \vec{a})\vec{v}}{v \|\vec{v} \times \vec{a}\|}, \quad a = \frac{\vec{a} \cdot \vec{v}}{v}, \quad \kappa = \frac{\|\vec{v} \times \vec{a}\|}{v^3}. \quad (4.1)$$

Ejemplo Si aplicamos las fórmulas anteriores a la parábola $\vec{r}(t) = (t, t^2)$ obtenemos:

$$\vec{v} = (1, 2t), \quad \vec{a} = (0, 2), \quad v = \sqrt{1 + 4t^2}, \quad \vec{\tau} = \left(\frac{1}{\sqrt{1 + 4t^2}}, \frac{2t}{\sqrt{1 + 4t^2}} \right),$$

$$\vec{n} = \left(-\frac{2t}{\sqrt{1 + 4t^2}}, \frac{1}{\sqrt{1 + 4t^2}} \right), \quad a = \frac{4t}{\sqrt{1 + 4t^2}}, \quad \kappa = \frac{2}{(1 + 4t^2)^{3/2}}. \quad \blacksquare$$

Se define el *radio de curvatura* de una curva paramétrica \vec{r} en un punto t como $\rho(t) = 1/\kappa(t)$, que es el radio de la circunferencia que tiene curvatura $\kappa(t)$. Más aún, se define la *circunferencia osculatriz* de la curva en el punto t como la circunferencia de centro $\vec{r} + \rho\vec{n}$ y radio ρ .



La figura muestra la circunferencia osculatriz en un punto de la parábola del ejemplo anterior. Es fácil ver que, tal y como muestra la figura, coincide con la curva en el punto sobre el que se calcula, tiene la misma recta tangente y, de hecho, tiene la misma curvatura en dicho punto. La figura de la derecha muestra la consecuencia geométrica de este hecho: hay un tramo en el que la curva y la recta tangente se distinguen claramente, pero la curva y la circunferencia siguen siendo indistinguibles.

Si la recta tangente es la trayectoria que seguiría el objeto en caso de no tener aceleración normal, la circunferencia osculatriz es la trayectoria que seguiría el objeto en caso de que el módulo de su aceleración normal fuera constante.

Ahora definimos el *vector binormal* como

$$\vec{b} = \vec{\tau} \times \vec{n}.$$

Explícitamente, teniendo en cuenta que $\vec{v} \times \vec{v} = \vec{0}$, tenemos que

$$\vec{b} = \frac{\vec{v} \times \vec{a}}{\|\vec{v} \times \vec{a}\|}.$$

De este modo, en cada instante t , los vectores $(\vec{\tau}, \vec{n}, \vec{b})$ son unitarios y ortogonales dos a dos, luego determinan un sistema de referencia de origen en \vec{r} . Cualquier vector puede expresarse en la forma $\alpha\vec{\tau} + \beta\vec{n} + \gamma\vec{b}$, para ciertas coordenadas (α, β, γ) . Esto vale en particular para su derivada:

$$\vec{b}' = \alpha\vec{\tau}' + \beta\vec{n}' + \gamma\vec{b}',$$

donde α, β, γ son funciones de t . Ahora bien, como $\vec{b} \cdot \vec{b} = 1$, al derivar queda $\vec{b} \cdot \vec{b}' = 0$, luego multiplicando la igualdad anterior queda $0 = \gamma$.

Similarmente, como $\vec{b} \cdot \vec{\tau} = 0$, al derivar obtenemos que $\vec{b}' \cdot \vec{\tau} + \vec{b} \cdot \vec{\tau}' = \vec{0}$, y así

$$\alpha = \vec{b}' \cdot \vec{\tau} = -\vec{b} \cdot \vec{\tau}' = 0,$$

pues $\vec{\tau}'$ tiene la dirección de \vec{n} , luego es ortogonal a \vec{b} .

Concluimos que \vec{b}' tiene siempre la dirección de \vec{n} . Definimos la *torsión* como

$$\tau = -\frac{\vec{b}' \cdot \vec{n}}{v},$$

de modo que, en la expresión general para \vec{b}' es $\beta = \vec{b}' \cdot \vec{n} = v\tau$ y, en definitiva:

$$\vec{b}' = -v\tau\vec{n}.$$

Como en el caso de la curvatura, hemos introducido la v en la definición de torsión para que ésta sea intrínseca a la curva, es decir, independiente de la parametrización.

En efecto, si consideramos una curva $\vec{r}(t)$ y una reparametrización arbitraria $\vec{r}(u) = \vec{r}(t(u))$, tenemos que

$$\vec{v}(u) = \vec{r}'(t(u)) \frac{dt}{du}, \quad v(u) = v(t(u)) \left| \frac{dt}{du} \right|, \quad \vec{\tau}(u) = \pm \vec{\tau}'(t(u)),$$

$$\vec{\tau}'(u) = \pm \vec{\tau}''(t(u)) \frac{dt}{du} = \vec{\tau}''(t(u)) \left| \frac{dt}{du} \right|, \quad \|\vec{\tau}'(u)\| = \|\vec{\tau}''(t(u))\| \left| \frac{dt}{du} \right|,$$

$$\vec{n}(u) = \vec{n}(t(u)), \quad \vec{b}(u) = \pm \vec{\tau}'(t(u)) \times \vec{n}(t(u)) = \pm \vec{b}'(t(u)),$$

$$\vec{b}'(u) = \pm \vec{b}''(t(u)) \frac{dt}{du} = \vec{b}''(t(u)) \left| \frac{dt}{du} \right|,$$

y finalmente:

$$\tau(u) = -\frac{\vec{b}'(u) \cdot \vec{n}(u)}{v(u)} = -\frac{\vec{b}''(t(u)) \left| \frac{dt}{du} \right| \cdot \vec{n}(t(u))}{v(t(u)) \left| \frac{dt}{du} \right|} = \tau(t(u)).$$

Con esto casi tenemos demostrado el teorema siguiente:

Teorema 4.1 (Fórmulas de Frenet) Si \vec{r} es una curva paramétrica con velocidad y aceleración no nulas, entonces

$$\frac{d\vec{\tau}}{dt} = v\kappa\vec{n}, \quad \frac{d\vec{n}}{dt} = -v\kappa\vec{\tau} + v\tau\vec{b}, \quad \frac{d\vec{b}}{dt} = -v\tau\vec{n}.$$

DEMOSTRACIÓN: Sólo falta calcular la derivada de \vec{n} . Como en el caso de \vec{b}' , podemos expresar:

$$\vec{n}' = \alpha\vec{\tau} + \beta\vec{n} + \gamma\vec{b},$$

pero $\vec{n} \cdot \vec{n} = 1$ implica que $\beta = \vec{n}' \cdot \vec{n} = 0$. Por otro lado, como $\vec{n} \cdot \vec{\tau} = 0$, derivando queda que

$$\alpha = \vec{n}' \cdot \vec{\tau} = -\vec{n} \cdot \vec{\tau}' = -\vec{n} \cdot (v\kappa\vec{n}) = -v\kappa.$$

Similarmente, $\gamma = \vec{n}' \cdot \vec{b} = -\vec{n} \cdot \vec{b}' = \vec{n} \cdot (v\tau\vec{n}) = v\tau$. ■

Es posible dar una fórmula explícita para la torsión como la que hemos obtenido para la curvatura. Para ello observamos en primer lugar (es una comprobación rutinaria) que

$$\frac{d(\vec{u} \times \vec{v})}{dt} = \frac{d\vec{u}}{dt} \times \vec{v} + \vec{u} \times \frac{d\vec{v}}{dt}.$$

Derivando la expresión

$$\vec{b} = \frac{\vec{v} \times \vec{a}}{\|\vec{v} \times \vec{a}\|},$$

obtenemos

$$\frac{d\vec{b}}{dt} = \frac{d\|\vec{v} \times \vec{a}\|^{-1}}{dt} \vec{v} \times \vec{a} + \|\vec{v} \times \vec{a}\|^{-1} (\vec{a} \times \vec{a}' + \vec{v} \times \vec{a}')$$

Ahora observamos que $\vec{a} \times \vec{a} = \vec{0}$ y, multiplicando por \vec{n} queda:

$$-\tau v = \vec{n} \cdot \frac{d\vec{b}}{dt} = \frac{v^2 \vec{a} - (\vec{v} \cdot \vec{a})\vec{v}}{v \|\vec{v} \times \vec{a}\|} \left(\frac{d\|\vec{v} \times \vec{a}\|^{-1}}{dt} \vec{v} \times \vec{a} + \|\vec{v} \times \vec{a}\|^{-1} \vec{v} \times \vec{a}' \right)$$

Pero \vec{v} es ortogonal tanto a $\vec{v} \times \vec{a}$ como a $\vec{v} \times \vec{a}'$, luego podemos eliminar el término $(\vec{v} \cdot \vec{a})\vec{v}$, y \vec{a} es ortogonal a $\vec{v} \times \vec{a}$, luego también podemos eliminar el primer sumando del paréntesis. Queda entonces que

$$-\tau v = \frac{v\vec{a}}{\|\vec{v} \times \vec{a}\|^2} \vec{v} \times \vec{a}',$$

luego, en términos del producto mixto A.7,

$$\tau = -\frac{\vec{a} \cdot (\vec{v} \times \vec{a}')}{\|\vec{v} \times \vec{a}\|^2} = \frac{(\vec{v}, \vec{a}, \vec{a}')}{\|\vec{v} \times \vec{a}\|^2}.$$

Las curvas planas se caracterizan por que su torsión es nula:

Teorema 4.2 Una curva $\vec{r}(t)$ está contenida en un plano si y sólo si su torsión es nula.

DEMOSTRACIÓN: Si la torsión es nula, entonces el vector binormal \vec{b} es constante. Vamos a probar que $(\vec{r}(t) - \vec{r}(t_0)) \cdot \vec{b} = 0$, lo que significa que $\vec{r}(t)$ está en el plano que pasa por $\vec{r}(t_0)$ con vector director \vec{b} . En efecto,

$$((\vec{r}(t) - \vec{r}(t_0)) \cdot \vec{b})' = \vec{v}(t) \cdot \vec{b} = 0,$$

luego $(\vec{r}'(t) - \vec{r}'(t_0)) \cdot \vec{b} = c$ es constante, luego

$$0 = \vec{v}(t_0) \cdot \vec{b} = \lim_{h \rightarrow 0} \frac{\vec{r}'(t_0 + h) - \vec{r}'(t_0)}{h} \cdot \vec{b} = \lim_{h \rightarrow 0} \frac{c}{h},$$

lo cual sólo es posible si $c = 0$.

Recíprocamente, si la curva está contenida en un plano de vector director \vec{u} , que podemos tomar de norma 1, entonces

$$(\vec{r}'(t) - \vec{r}'(t_0)) \cdot \vec{u} = 0,$$

luego, derivando, $\vec{v}(t) \cdot \vec{u} = 0$, luego $\vec{\tau} \cdot \vec{u} = 0$ y, volviendo a derivar, $\vec{\tau}' \cdot \vec{u} = 0$, y esto implica que $\vec{n} \cdot \vec{u} = 0$, luego $\vec{\tau}$ y \vec{n} son ortogonales a \vec{u} , luego $\vec{b} = \vec{\tau} \times \vec{n} = \pm \vec{u}$ es constante, luego la torsión es nula. ■

Ejemplo: La hélice Consideremos la hélice parametrizada por el arco

$$\vec{r}(s) = (a \cos(s/c), a \sin(s/c), bs/c), \quad c = \sqrt{a^2 + b^2}.$$

Tenemos que

$$\vec{\tau} = \vec{v}' = (-a/c \sin(s/c), a/c \cos(s/c), b/c),$$

$$v = \sqrt{a^2/c^2 + b^2/c^2} = 1,$$

luego

$$\vec{\tau}' = \frac{a}{c^2} (-\cos(s/c), -\sin(s/c), 0),$$

de donde $\kappa = \|\vec{\tau}'\| = a/c^2$ y $\vec{n} = (-\cos(s/c), -\sin(s/c), 0)$. A su vez

$$\vec{b} = \vec{\tau} \times \vec{n} = \frac{1}{c} (b \sin(s/c), -b \cos(s/c), a),$$

luego

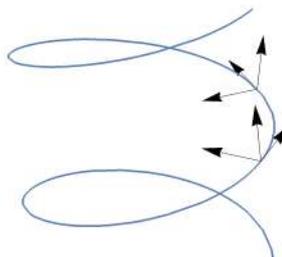
$$\vec{b}' = \frac{1}{c^2} (b \cos(s/c), b \sin(s/c), 0) = -\frac{b}{c^2} \vec{n},$$

luego

$$\tau = \frac{b}{a^2 + b^2}.$$

Así pues, la hélice tiene curvatura y torsión constante. La torsión es positiva cuando para ascender hay que girar en sentido positivo y es negativa cuando para ascender hay que girar en sentido negativo. ■

El teorema siguiente muestra que la curvatura y la torsión determinan completamente la forma de una curva:



Teorema 4.3 Si dos curvas $\vec{r}_1, \vec{r}_2 : I \rightarrow \mathbb{R}^3$ parametrizadas por el arco tienen las mismas funciones de curvatura y torsión y, para un cierto $s_0 \in I$, cumplen $\vec{r}_1(s_0) = \vec{r}_2(s_0)$, $\vec{\tau}_1(s_0) = \vec{\tau}_2(s_0)$, $\vec{n}_1(s_0) = \vec{n}_2(s_0)$ y $\vec{b}_1(s_0) = \vec{b}_2(s_0)$, entonces $\vec{r}_1 = \vec{r}_2$, es decir, ambas curvas son la misma.

DEMOSTRACIÓN: Basta observar que

$$\begin{aligned} & \frac{1}{2} \frac{d}{ds} (\|\vec{\tau}_1 - \vec{\tau}_2\|^2 + \|\vec{n}_1 - \vec{n}_2\|^2 + \|\vec{b}_1 - \vec{b}_2\|^2) = \\ & \frac{1}{2} \frac{d}{ds} ((\vec{\tau}_1 - \vec{\tau}_2)(\vec{\tau}_1 - \vec{\tau}_2) + (\vec{n}_1 - \vec{n}_2)(\vec{n}_1 - \vec{n}_2) + (\vec{b}_1 - \vec{b}_2)(\vec{b}_1 - \vec{b}_2)) = \\ & (\vec{\tau}_1 - \vec{\tau}_2)(\vec{\tau}_1' - \vec{\tau}_2') + (\vec{n}_1 - \vec{n}_2)(\vec{n}_1' - \vec{n}_2') + (\vec{b}_1 - \vec{b}_2)(\vec{b}_1' - \vec{b}_2'), \end{aligned}$$

y aplicando las fórmulas de Frenet (teniendo en cuenta que, al estar parametrizadas por el arco, las curvas cumplen $v_1 = v_2 = 1$), queda:

$$\kappa(\vec{\tau}_1 - \vec{\tau}_2)(\vec{n}_1 - \vec{n}_2) - \kappa(\vec{n}_1 - \vec{n}_2)(\vec{\tau}_1 - \vec{\tau}_2) + \tau(\vec{n}_1 - \vec{n}_2)(\vec{b}_1 - \vec{b}_2) - \tau(\vec{b}_1 - \vec{b}_2)(\vec{n}_1 - \vec{n}_2),$$

es decir,

$$\frac{d}{ds} (\|\vec{\tau}_1 - \vec{\tau}_2\|^2 + \|\vec{n}_1 - \vec{n}_2\|^2 + \|\vec{b}_1 - \vec{b}_2\|^2) = 0.$$

Por lo tanto, la función $\|\vec{\tau}_1 - \vec{\tau}_2\|^2 + \|\vec{n}_1 - \vec{n}_2\|^2 + \|\vec{b}_1 - \vec{b}_2\|^2$ es constante y, como en s_0 se anula, concluimos que

$$\|\vec{\tau}_1 - \vec{\tau}_2\|^2 + \|\vec{n}_1 - \vec{n}_2\|^2 + \|\vec{b}_1 - \vec{b}_2\|^2 = 0.$$

Esto implica que $\vec{\tau}_1 = \vec{\tau}_2$, $\vec{n}_1 = \vec{n}_2$ y $\vec{b}_1 = \vec{b}_2$. Pero, teniendo en cuenta de nuevo que las curvas están parametrizadas por el arco, la primera igualdad es $\vec{r}_1' = \vec{r}_2'$, o también $(\vec{r}_1 - \vec{r}_2)' = 0$, luego $\vec{r}_1 - \vec{r}_2$ es constante y, teniendo en cuenta que $\vec{r}_1(s_0) - \vec{r}_2(s_0) = \vec{0}$, tiene que ser $\vec{r}_1 - \vec{r}_2 = \vec{0}$. ■

Nota Observemos que si dos curvas \vec{r}_1 y \vec{r}_2 parametrizadas por el arco tienen las mismas funciones de curvatura y torsión, siempre podemos mover \vec{r}_2 para que se cumplan las condiciones $\vec{r}_1(s_0) = \vec{r}_2(s_0)$, $\vec{\tau}_1(s_0) = \vec{\tau}_2(s_0)$, $\vec{n}_1(s_0) = \vec{n}_2(s_0)$ y $\vec{b}_1(s_0) = \vec{b}_2(s_0)$, y el teorema anterior nos da que $\vec{r}_1 = \vec{r}_2$. En otras palabras, si dos curvas parametrizadas por el arco tienen las mismas funciones de curvatura y torsión, entonces son iguales en el sentido geométrico de que una puede moverse hasta superponerse con la otra. ■

Vamos a discutir ahora el concepto de velocidad angular, para lo cual empecemos considerando algunos casos particulares:

El movimiento circular uniforme Consideremos un objeto puntual cuya trayectoria es

$$\vec{r}(t) = (r \cos \omega t, r \sin \omega t),$$

para ciertas constantes r y ω , cuya interpretación es clara: el objeto está recorriendo una circunferencia de radio r y ω son los radianes que gira cada segundo.

Por ejemplo, si $\omega = 2\pi$ tenemos que el objeto gira 2π radianes (es decir, da una vuelta completa) en un segundo, mientras que si $\omega = 1$ el objeto avanza 1 radian cada segundo.

En general, se dice que el objeto sigue un *movimiento circular uniforme* de radio r y *velocidad angular* ω .

El *periodo* de rotación $T = 2\pi/\omega$ es el tiempo que tarda el objeto en dar una vuelta completa, y la *frecuencia* de rotación $\nu = 1/T = \omega/2\pi$ es el número de vueltas que da en un segundo. Observemos que

$$\vec{v} = (-r\omega \operatorname{sen} \omega t, r\omega \operatorname{cos} \omega t), \quad v = r|\omega|.$$

La última relación tiene una interpretación clara: por la propia definición de “radián”, recorrer un radian sobre una circunferencia de r metros de radio equivale a recorrer r metros sobre la circunferencia, luego $|\omega|$ radianes por segundo equivalen a $r|\omega|$ metros por segundo. Volviendo a derivar:

$$\vec{a} = (-r\omega^2 \operatorname{cos} \omega t, -r\omega^2 \operatorname{sen} \omega t).$$

Vemos así que la aceleración de un objeto en movimiento circular uniforme apunta en todo momento hacia el centro del giro y su módulo es $a = \omega^2 r$. Como era de esperar, la aceleración coincide con la aceleración normal, ya que, como v no varía, no hay aceleración tangencial. ■

Movimiento circular Un poco más en general, si un cuerpo se mueve sobre una circunferencia de radio r , su trayectoria será de la forma

$$\vec{r}(t) = (r \operatorname{cos} \theta(t), r \operatorname{sen} \theta(t)),$$

de modo que, definiendo ahora la *velocidad angular* como $\omega = \theta'$, tenemos igualmente

$$\vec{v} = (-r\omega \operatorname{sen} \theta, r\omega \operatorname{cos} \theta), \quad v = r|\omega|.$$

Notemos que el movimiento circular uniforme es el caso particular que resulta de tomar $\theta(t) = \omega t + \theta_0$.

Ahora la velocidad angular $\omega = \theta'$ se interpreta como cualquier derivada: indica la variación del ángulo θ por unidad de tiempo en un instante dado, entendiendo que si, por ejemplo, $\omega = 3 \operatorname{rad/s}$, esto no significa que el objeto vaya a recorrer 3 radianes en un segundo, sino que, *en caso de mantener la misma velocidad angular*, recorrería 3 radianes en un segundo. Volviendo a derivar:

$$\vec{a} = (-r\alpha \operatorname{sen} \theta, r\alpha \operatorname{cos} \theta) - (r\omega^2 \operatorname{cos} \theta, r\omega^2 \operatorname{sen} \theta),$$

donde $\alpha = \omega' = \theta''$ es la *aceleración angular*, de modo que la aceleración tangencial tiene módulo $a_t = |\alpha|r$ y la aceleración normal $a_n = \omega^2 r$, como en el caso uniforme. ■

La velocidad angular Consideremos una situación más general en la que un objeto puntual se mueve en un plano sin pasar nunca por un cierto punto que tomamos como origen del sistema de referencia. Entonces su trayectoria puede expresarse igualmente como

$$\vec{r} = (r \cos \theta, r \sin \theta),$$

donde ahora, tanto r como θ son funciones de t . (El caso precedente es el caso particular que resulta de tomar r constante.)

Imaginemos un observador situado en el origen de coordenadas que ve el objeto, pero no es capaz de medir la distancia r a la que se encuentra en cada momento (por ejemplo, podemos ver un coche moviéndose a lo lejos, de modo que no sabemos a qué distancia está, pero sí que podemos determinar perfectamente con qué ángulo θ lo vemos respecto de unos ejes prefijados).

Podemos definir igualmente la *velocidad angular* como $\omega = \theta'$, que expresa igualmente la variación del ángulo θ por unidad de tiempo que puede constatar el observador, sin entrar en la posible variación de la distancia r . Ahora tenemos

$$\vec{v} = r'(\cos \theta, \sin \theta) + r\omega(-\sin \theta, \cos \theta) = r'\vec{u}_r + r\omega\vec{u}_t,$$

donde \vec{u}_r y \vec{u}_t son, respectivamente, el *vector unitario radial* y el *vector unitario transversal*. Así, hemos descompuesto el vector velocidad en una *velocidad radial* $\vec{v}_r = r'\vec{u}_r$, que mide la variación de la distancia del objeto, y una *velocidad transversal* $\vec{v}_t = r\omega\vec{u}_t$, que es la velocidad que tendría el objeto en un instante dado si se estuviera moviendo en una trayectoria circular con la misma velocidad angular (es decir, si fuera $r' = 0$ y, por consiguiente, la velocidad radial fuera nula y la velocidad coincidiera con la velocidad transversal).

En otras palabras, si el observador supiera que el objeto se encuentra a una distancia r en un momento dado, pero no fuera capaz de apreciar la variación de r y supusiera que el movimiento es circular, entonces, al medir una velocidad angular ω concluiría que la velocidad sobre la trayectoria es $v = r|\omega|$ y, más precisamente, que la velocidad es $\vec{v} = \vec{v}_t$.

Veamos ahora que ω puede calcularse sin necesidad de hacer referencia a θ . Para ello observamos que

$$\|\vec{r} \times \vec{v}\| = \|\vec{r} \times \vec{v}_t\| = r^2|\omega|,$$

ya que \vec{r} y \vec{v}_t son ortogonales. Esto nos lleva a definir el *vector velocidad angular* como

$$\vec{\omega} = \frac{1}{r^2} \vec{r} \times \vec{v},$$

que es ortogonal al plano del movimiento y su módulo es $|\omega|$. Más aún, el signo de ω es positivo si el sentido de giro es antihorario y negativo en caso contrario, y esta información está también en el vector $\vec{\omega}$, pues, por la regla de la mano derecha, $\vec{\omega}$ apunta hacia el semiespacio determinado por el plano de giro desde el cual éste se ve en sentido positivo, luego conociendo $\vec{\omega}$ sabemos en qué sentido está girando el objeto.

Más precisamente, a partir de $\vec{\omega}$ podemos recuperar la velocidad transversal como

$$\vec{v}_t = \vec{\omega} \times \vec{r}.$$

En efecto, ambos miembros tiene módulo $r|\omega|$, ambos son ortogonales a \vec{r} y a $\vec{\omega}$, luego tienen la misma dirección, y la regla de la mano derecha (véase la figura del ejemplo siguiente) muestra que ambos tienen el mismo sentido, luego son iguales.

Ejemplo Consideremos la trayectoria elíptica

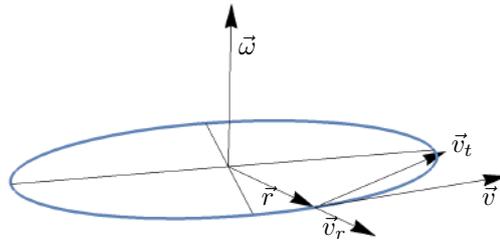
$$\vec{r} = (2 \cos t, \sin t).$$

Se cumple que $r = \sqrt{1 + 3 \cos^2 t}$, $\vec{v}(t) = (-2 \sin t, \cos t)$,

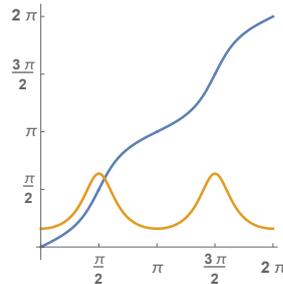
$$\vec{r} \times \vec{v} = \begin{vmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ 2 \cos t & \sin t & 0 \\ -2 \sin t & \cos t & 0 \end{vmatrix} = (0, 0, 2),$$

luego

$$\vec{\omega} = (0, 0, \frac{2}{1 + 3 \cos^2 t}).$$



Vemos que la velocidad angular no es constante. La figura siguiente muestra las gráficas de ω y θ :

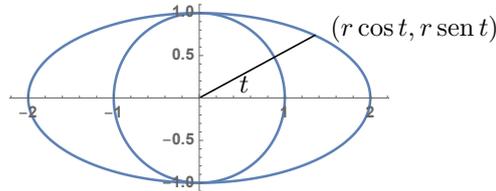


Desde el origen de coordenadas se ve que la partícula gira más rápidamente cuando está más cerca y más lentamente cuando está más lejos.

Podemos reparametrizar la elipse para recorrerla con velocidad angular constante. Para ello observamos en primer lugar que la ecuación de la elipse es

$$\frac{x^2}{4} + \frac{y^2}{1} = 1.$$

El punto de la elipse que forma un ángulo t con el semieje de abscisas positivas será de la forma $(r \cos t, r \sin t)$, para cierto r que podemos calcular exigiendo que el punto cumpla la ecuación de la elipse:



$$\frac{r^2 \cos^2 t}{4} + r^2 \sin^2 t = 1,$$

de donde

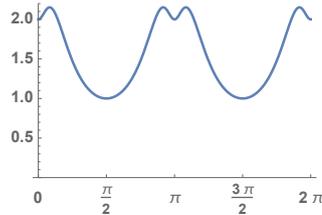
$$r^2 = \frac{4}{\cos^2 t + 4 \sin^2 t} = \frac{4}{1 + 3 \sin^2 t},$$

luego

$$\vec{r}(t) = \left(\frac{2 \cos t}{\sqrt{1 + 3 \sin^2 t}}, \frac{2 \sin t}{\sqrt{1 + 3 \sin^2 t}} \right).$$

El lector puede comprobar que, a partir de esta parametrización, se obtiene $\vec{\omega} = (0, 0, 1)$, pero el cálculo es muy farragoso y sabemos *a priori* que éste tiene que ser el resultado, por construcción.

Si una partícula se mueve siguiendo esta parametrización y un observador situado en el origen de coordenadas no es capaz de medir la distancia a la que se encuentra, es decir, si no puede distinguir si la partícula se encuentra en $(r \cos t, r \sin t)$ o en $(\cos t, \sin t)$, lo que observa será compatible con que la partícula se esté moviendo a lo largo de una circunferencia con velocidad constante. En realidad no es así, sino que la velocidad varía para que la velocidad angular no varíe. Ésta es la gráfica del módulo de la velocidad v :



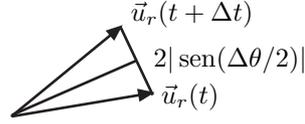
Finalmente, vamos a considerar el caso general de un objeto puntual que se mueve por el espacio sin pasar por un punto que tomamos como origen de coordenadas. Entonces podemos definir el módulo $r = \|\vec{r}\|$ y el vector radial unitario

$$\vec{u}_r = \frac{1}{r} \vec{r}.$$

Imaginemos, como antes, que un observador situado en el origen de coordenadas ve el objeto, pero no puede medir la distancia r a la que se encuentra.

Lo que sí que puede medir es el ángulo que forman los vectores $\vec{r}(t)$ y $\vec{r}(t + \Delta t)$, para un instante inicial fijo t y un incremento de tiempo Δt . Dicho ángulo es el mismo que forman los vectores $\vec{u}_r(t)$ y $\vec{u}_r(t + \Delta t)$. Como estos vectores son unitarios, tenemos que

$$\|\vec{u}_r(t + \Delta t) - \vec{u}_r(t)\| = 2 \left| \sin \frac{\Delta\theta}{2} \right|.$$



luego

$$\left\| \frac{\vec{u}_r(t + \Delta t) - \vec{u}_r(t)}{\Delta t} \right\| = \frac{\|\vec{u}_r(t + \Delta t) - \vec{u}_r(t)\|}{|\Delta t|} = \left| \frac{\sin \frac{\Delta\theta}{2}}{\Delta\theta/2} \right| \frac{\Delta\theta}{|\Delta t|}.$$

Es claro que $\Delta\theta \rightarrow 0$ cuando $\Delta t \rightarrow 0$, luego

$$\omega = \lim_{\Delta t \rightarrow 0} \frac{\Delta\theta}{|\Delta t|} = \|\vec{u}'_r(t)\|.$$

(Notemos que el hecho de que \vec{u}_r sea derivable implica la existencia del límite.) Con esto llegamos a la definición general de velocidad angular:

Definición 4.4 La *velocidad angular* de un objeto puntual de trayectoria \vec{r} en cada instante t (respecto del origen de coordenadas) es $\omega = \|\vec{u}'_r\|$.

Acabamos de probar que ω mide la variación del ángulo con el que se observa el móvil. Más precisamente, si Δt es suficientemente pequeño, el ángulo $\Delta\theta$ que forman los vectores $\vec{r}(t)$ y $\vec{r}(t + \Delta t)$ es aproximadamente $\Delta\theta \approx \omega |\Delta t|$ y la aproximación es mejor cuanto menor es Δt .

Vamos a probar que ω así definida es igual a $|\theta'|$ en los ejemplos precedentes, es decir, es el valor absoluto de la velocidad angular escalar que habíamos definido previamente. Para ello, siguiendo en el caso general, observamos que al derivar la relación $\vec{r} = r \vec{u}_r$ obtenemos

$$\vec{v} = r' \vec{u}_r + r \vec{u}'_r$$

y derivando la relación $\vec{u}_r \cdot \vec{u}_r = 1$ obtenemos que \vec{u}'_r es ortogonal a \vec{u}_r .

Así hemos descompuesto la velocidad \vec{v} en una *velocidad radial* $\vec{v}_r = r' \vec{u}_r$ y una *velocidad transversal* $\vec{v}_t = r \vec{u}'_r$, que es ortogonal a \vec{r} , tiene módulo $r\omega$. Es claro que, al igual que en el caso del movimiento plano, la velocidad radial mide la variación de la distancia del objeto, y para estudiar la velocidad transversal observamos que el vector $\vec{r} \times \vec{v} = \vec{r} \times \vec{v}_t$ tiene módulo $\|\vec{r} \times \vec{v}\| = r^2\omega$.

Definimos el *vector velocidad angular*

$$\vec{\omega} = \frac{1}{r^2} \vec{r} \times \vec{v}.$$

Así $\|\vec{\omega}\| = \omega$, y con esto ya tenemos probado que, en el caso de un movimiento plano, $\vec{\omega}$ es el mismo vector que ya teníamos definido, y que $\omega = |\theta'|$, como habíamos afirmado.

En general se cumple igualmente que $\vec{v}_t = \vec{\omega} \times \vec{r}$, pues, como antes, ambos vectores tienen módulo $r\omega$, ambos son ortogonales a \vec{r} y a $\vec{\omega}$, y la regla de la mano derecha nos da que ambos tienen el mismo sentido.

Si $\vec{\omega}$ mantiene su dirección constante (aunque varíe su módulo), entonces, como $\vec{r} \cdot \vec{\omega} = 0$, esto significa que \vec{r} está siempre en el plano que pasa por el origen y es perpendicular a la recta sobre la que varía $\vec{\omega}$. Por eso podemos decir que $\vec{\omega}$ determina el plano de giro en cada instante (aunque éste pueda estar cambiando continuamente).

Más aún, siguiendo con el supuesto de que la dirección de $\vec{\omega}$ sea constante, tomando un sistema de referencia que haga que apunte en la dirección del eje Z , la trayectoria (que ya hemos visto que es plana), puede expresarse como

$$\vec{r} = (r \cos \theta, r \sin \theta, 0)$$

y hemos visto que $\vec{v} = \dot{r} \vec{u}'_r + r \dot{\theta} \vec{u}'_t$ y entonces

$$\vec{\omega} = \vec{r} \times r \dot{\theta} \vec{u}'_t = (0, 0, \dot{\theta}),$$

luego el vector $\vec{\omega}$ contiene toda la información sobre la forma en que el objeto está girando alrededor del origen de coordenadas (plano de giro, sentido de giro y velocidad angular), independientemente de lo que el objeto se acerque o se aleje del origen.

En general, aunque $\vec{\omega}$ no sea constante, podemos interpretar que su dirección indica la dirección del eje alrededor del cual está girando el objeto en un instante t , cuyo sentido determina el sentido de giro (es positivo visto desde el semiplano hacia el que apunta $\vec{\omega}$) y cuyo módulo ω indica —según hemos visto— la velocidad angular del giro en cada instante, sin tener en cuenta la posibilidad de que el objeto esté acercándose o alejándose del origen, sobre lo cual $\vec{\omega}$ no aporta ninguna información.

Derivando la definición de $\vec{\omega}$ podemos obtener una expresión para la *aceleración angular*

$$\vec{\alpha} = \dot{\vec{\omega}} = -\frac{2r'}{r^3} \vec{r} \times \vec{v} + \frac{1}{r^2} (\vec{v} \times \vec{v} + \vec{r} \times \vec{a}),$$

de donde

$$\vec{\alpha} = -\frac{2r'}{r} \vec{\omega} + \frac{1}{r^2} \vec{r} \times \vec{a}.$$

En el caso particular en que r es constante, es decir, si la trayectoria está contenida en una esfera de radio r , esta fórmula se reduce a

$$\vec{\alpha} = \frac{1}{r^2} \vec{r} \times \vec{a}.$$

Además en este caso la velocidad transversal es simplemente la velocidad, por lo que $\vec{v} = \vec{\omega} \times \vec{r}$.

Si además la trayectoria es plana (es decir, circular), sabemos que respecto de un sistema de referencia adecuado, $\vec{\omega} = (0, 0, \dot{\theta})$, luego $\vec{\alpha} = (0, 0, \ddot{\theta})$ y así el módulo de $\vec{\alpha}$ indica la velocidad con la que varía ω , aumentando si $\vec{\alpha}$ tiene el sentido de $\vec{\omega}$ y disminuyendo si el sentido es el opuesto. ■

4.2 Las leyes de Newton

Nada de cuanto hemos dicho en la sección anterior nos permite predecir cómo se moverá un objeto en unas condiciones dadas. En esta sección presentamos los principios dinámicos básicos que nos permitirán aplicar el cálculo diferencial a este tipo de problemas. Se trata de las tres leyes de Newton que ya hemos discutido en la introducción, la primera de las cuales es el principio de inercia:

Primera ley de Newton *Si un cuerpo está libre de toda influencia externa, su velocidad permanece constante, es decir, o permanece en reposo, o se mueve en línea recta con velocidad constante.*

En otras palabras, un cuerpo libre de toda influencia externa se mueve sin aceleración o, recíprocamente, si un cuerpo experimenta una aceleración, ésta tiene que ser el efecto de una influencia externa.

Fuerzas Ahora bien, ¿en qué consisten exactamente esas “influencias externas” que causan aceleraciones? Es obvio que si le damos una patada a una piedra que está en reposo y ésta empieza a moverse, la “influencia externa” que ha causado su aceleración ha sido la patada que ha recibido, pero existen “influencias externas” más sutiles. Por ejemplo, si sostengo una piedra en mi mano y la suelto sin empujarla, ésta deja de estar en reposo y cae. ¿Cuál es ahora la “influencia externa” que la ha acelerado, si mi mano no la ha empujado? La respuesta es que la ha acelerado la gravedad terrestre. De algún modo, la piedra “sabe” que ahí, a cierta distancia, está la Tierra y se siente atraída por ella, igual que, a su vez, la Tierra “sabe” que ahí, a cierta distancia, está el Sol.

Sucede que las causas capaces de acelerar un cuerpo son realmente muy pocas: apreciables a nivel macroscópico —esto es, descartando algunas que sólo afectan a las partículas subatómicas— sólo hay tres: la fuerza gravitatoria y otras dos muy relacionadas entre sí: la fuerza eléctrica y la fuerza magnética.

Cuando dejamos caer una piedra, ésta se mueve porque experimenta la fuerza de atracción gravitatoria de la Tierra, cuando le pegamos una patada a una piedra, ésta se mueve porque los electrones de los átomos de la punta de mi zapato ejercen una fuerza repulsiva sobre los electrones de la superficie de la piedra, y así, la inmensa mayoría de las aceleraciones que percibimos en los objetos son causadas por la fuerza gravitatoria o la fuerza eléctrica (y sólo unas pocas por la fuerza magnética, como cuando un imán atrae una pieza metálica).

A la hora de plasmar matemáticamente estas ideas, una fuerza se representa matemáticamente como un vector. Puesto que las fuerzas son las causas de las aceleraciones, podríamos pensar que, matemáticamente, un vector de fuerza que actúa sobre un objeto puede identificarse con su vector de aceleración, pero esto es falso por dos razones:

Por una parte, la aceleración que sufren en un momento dado un objeto es la que es, y no puede ser más que una, mientras que sobre un mismo objeto pueden actuar simultáneamente varias fuerzas. Por ejemplo, si un libro está en reposo sobre mi mesa, ello no se debe al principio de inercia, es decir, no es cierto que esté libre de toda influencia externa. Por el contrario, sobre él están

actuando dos fuerzas: la fuerza gravitatoria terrestre, que lo atrae hacia abajo, y la fuerza repulsiva que ejercen los electrones de la mesa sobre los electrones del libro, que lo empujan hacia arriba. Si el libro permanece en reposo ello se debe a que ambas fuerzas se cuantifican mediante vectores opuestos de suma nula, por lo que en la práctica es como si no se ejerciera ninguna fuerza sobre él. Más precisamente: no es cierto que una fuerza ejercida sobre un objeto se materialice siempre en una aceleración, sino que la aceleración del objeto dependerá de la suma de todas las fuerzas que actúan sobre él. Esto obliga a distinguir entre las fuerzas y la aceleración que provocan.

En segundo lugar, una fuerza pretende plasmar una causa física, la cual puede provocar aceleraciones distintas en cuerpos distintos. Por ejemplo, si le pegamos una patada a una piedra y después le pegamos una patada en las mismas condiciones a una pelota más ligera, la pelota experimentará una aceleración mayor que la piedra, por lo que tenemos dos fuerzas iguales que han causado aceleraciones diferentes.

La segunda ley de Newton afirma que las diferencias en la aceleración que una misma fuerza causa en objetos diferentes depende únicamente de una magnitud asociada a cada cuerpo, que es lo que se conoce como su *masa*:

Segunda ley de Newton *La suma \vec{F} de todas las fuerzas que actúan sobre un objeto en un instante dado es el producto de su masa m por la aceleración \vec{a} que experimenta en dicho instante: $\vec{F} = m\vec{a}$.*

Antes de poner ejemplos conviene precisar todo lo dicho hasta aquí con algunas observaciones sobre las unidades de medida:

Unidades de medida No es posible aplicar las matemáticas a la realidad física sin especificar un sistema de unidades de medida de las distintas magnitudes. Por ejemplo, en matemática pura es posible hablar de una “unidad de longitud” sin especificar cuánto mide exactamente, pero cualquier medición física es relativa a una unidad de longitud concreta que hay que especificar de un modo u otro. El sistema de unidades físicas más utilizado hoy en día es el *Sistema Internacional de Unidades*, fijado por la *Conferencia General de Pesas y Medidas*, que se reunió por primera vez en 1889.

De entre las unidades de todas las magnitudes que hemos considerado hasta ahora hay únicamente tres *unidades fundamentales*, es decir, unidades que tienen que ser definidas a partir de procesos físicos que puedan ser reproducidos en cualquier momento con garantías de que determinarán siempre la misma magnitud. Éstas son el *segundo* como unidad de tiempo, el *metro* como unidad de longitud y el *kilogramo* como unidad de masa.

La tabla siguiente recoge la definición antigua de cada una de estas unidades y la definición moderna fijada por la Conferencia General de Pesas y Medidas, que es mucho más técnica, pero permite realizar medidas de precisión.

segundo (s) (tradicional)	86 400-ava parte de un día solar, es decir, del tiempo que tarda el Sol en pasar dos veces consecutivas sobre un mismo meridiano.
(1967)	Duración de 9 192 631 770 oscilaciones de la radiación emitida en la transición entre dos niveles hiperfinos del estado fundamental del isótopo 133 del átomo de cesio a una temperatura de 0 K.
metro (m) (1792)	Diezmillonésima parte de la distancia que separa el polo norte de la línea del ecuador terrestre a través de la superficie terrestre.
(1983)	Distancia que recorre la luz en el vacío en un intervalo de 1/299 792 458 s.
kilogramo (kg) (1795)	Masa de un decímetro cúbico de agua pura en el punto de fusión del hielo.
(2018)	Unidad de masa que hace que la constante de Plank, es decir, la proporción entre la energía de un fotón y la frecuencia de su onda electromagnética asociada, sea igual a $h = 6.62607015 \cdot 10^{-34} \text{ kg}\cdot\text{m}^2\cdot\text{s}^{-1}$.

Admitiendo que sabemos cuánto es un segundo, un metro y un kilogramo (sea entendiendo las definiciones modernas, o aceptando las tradicionales como válidas para nuestros fines, o aceptando que nuestros instrumentos de medida funcionan correctamente), cualquier otra magnitud de las que hemos considerado hasta ahora se puede definir a partir de éstas.

Por ejemplo, la unidad para la velocidad en el Sistema Internacional es el *metro por segundo* (m/s), que es la velocidad de un móvil que recorre un metro cada segundo. Similarmente, la unidad para la aceleración es el *metro por segundo al cuadrado* (m/s²), que es la aceleración de un objeto que aumenta su velocidad en 1m/s cada segundo. Finalmente, la fuerza se mide en *newtons* (N = kg·m/s²), donde un newton es la fuerza que, cuando actúa sobre un cuerpo de 1 kg de masa, le produce una aceleración de 1 m/s².

Ejemplo: La gravedad terrestre Puede comprobarse que, en las cercanías de la superficie terrestre (a una escala en la que ésta puede considerarse plana), cada cuerpo es atraído por la Tierra con una fuerza dirigida hacia abajo y que es proporcional a su masa. Dicha fuerza es lo que se denomina el *peso* del objeto.

El factor de proporcionalidad se conoce como la *intensidad del campo gravitatorio terrestre* y se representa por la letra *g*.

Su valor depende ligeramente de la altitud, de la longitud y de la latitud, ya que la forma de la Tierra no es perfectamente esférica, y de las posibles irregularidades locales que pueda haber cerca de un punto concreto de su superficie (no es lo mismo estar en la cubierta de un barco en medio del océano que estar junto a una mina de plomo). No obstante, estas variaciones son despreciables en la mayoría de los contextos. Su valor puede determinarse experimentalmente con facilidad:

Ejemplo *Dejamos caer un objeto desde una altura de 5 m y llega al suelo al cabo de 1.01 s. Determinar la intensidad del campo gravitatorio terrestre.*

SOLUCIÓN: Tomando un sistema de referencia con origen en el suelo y en el que el eje Z sea vertical, es claro que sólo la tercera coordenada de todos los vectores implicados es relevante, así que podemos trabajar únicamente con ella. Si el objeto tiene masa m , la segunda ley de Newton dice que su aceleración en cada instante vendrá dada por la relación $-mg = ma$, luego $a = -g$. Así pues, vemos que un objeto en caída libre (es decir, sobre el que no actúa más fuerza que la gravedad) sufre una aceleración constante igual a g (dirigida hacia abajo) independientemente de su masa. Si su velocidad inicial es nula, su velocidad en cada instante t será

$$v = - \int_0^t g dt = -gt,$$

y si parte de una altura $h = 5$ m, su altura en un instante t será

$$h = 5 + \int_0^t -gt dt = 5 - \frac{1}{2}gt^2.$$

Como sabemos que $h(1.01) = 0$, se cumple que $5 - (1/2)g(1.01)^2 = 0$, de donde $g = 9.803 \text{ N/kg} = \text{m/s}^2$.

En la práctica el valor de g puede oscilar entre 9.764 y 9.834 N/kg, pero un valor promedio aceptado por convenio es $g = 9.80665 \text{ N/kg}$. Normalmente $g = 9.8$ es una aproximación razonable. ■

Observemos que, técnicamente, el peso de un cuerpo es una fuerza y se mide en newtons, al contrario que su masa, que se mide en kg. No obstante, el hecho de que ambos sean proporcionales hace que a menudo se diga que un cuerpo “pesa” 80 kg, cuando en realidad lo que se quiere decir es que ésa es su masa (y su peso es en realidad de $80 \cdot 9.8 = 784 \text{ N}$). La diferencia sólo es relevante si, por ejemplo, nos sumergimos en el agua —en cuyo caso la misma masa pesa menos— o si cambiamos de planeta. Por ejemplo, en la Luna la intensidad del campo gravitatorio es de 1.625 N/m, luego un astronauta de 80 kg pesa 130 N.

Así pues, un objeto sometido únicamente a la gravitación terrestre (en las proximidades de la superficie terrestre) experimenta una aceleración constante independientemente de su masa. En el caso concreto de la gravitación terrestre, respecto de un sistema de referencia cuya tercera coordenada sea la altura, la aceleración es $\vec{g} = (0, 0, -g)$. ■

Para completar los principios básicos de la mecánica newtoniana enunciamos el principio de acción y reacción:

Tercera ley de Newton *Cuando un cuerpo ejerce una fuerza sobre otro, éste ejerce a su vez una fuerza sobre el primero de la misma intensidad, pero dirigida en sentido contrario.*

Aquí es fundamental entender que una fuerza y su reacción correspondiente no se compensan, porque cada una se ejerce sobre un cuerpo distinto. Por

ejemplo, si dos personas con patines de ruedas en los pies están en reposo una ante la otra y una extiende sus brazos para empujar a la otra, ella misma se verá empujada hacia atrás con la misma fuerza que haya ejercido sobre la otra, aunque la segunda haya permanecido completamente pasiva.

Un ejemplo más cotidiano se produce siempre que un objeto está apoyado en otro, por ejemplo, un libro encima de una mesa. El libro está ejerciendo una fuerza sobre la mesa de intensidad igual a su peso y dirigida hacia abajo, y consecuentemente la mesa está ejerciendo sobre el libro una fuerza igual a su peso, pero dirigida hacia arriba. El resultado es que el libro permanece en reposo porque sobre él se están ejerciendo dos fuerzas opuestas: su peso y la reacción de la mesa.

Ejemplo Si dejamos caer una piedra de 1 kg desde una altura de 1 m, entonces la piedra experimenta una fuerza gravitatoria de 9.8 N que la hace caer hacia la Tierra, pero técnicamente no es cierto que la piedra caiga hacia la Tierra mientras ésta permanece en reposo, sino que la piedra también atrae a la Tierra con una fuerza de 9.8 N, por lo que en realidad la Tierra y la piedra corren una al encuentro de la otra. Ahora bien, teniendo en cuenta que la masa de la Tierra se estima en $M_T = 5.9736 \cdot 10^{24}$ kg, la aceleración que experimenta viene dada por la ecuación

$$M_T a_T = 9.8,$$

luego $a_T = 1.6405 \cdot 10^{-24}$ m/s². Si tomamos un sistema de referencia en el que la superficie de la Tierra está en $x = 0$ y la piedra en $x = 1$, se encontrarán en el instante t que cumpla:

$$\frac{1}{2} a_T t^2 = 1 - \frac{1}{2} g t^2,$$

luego

$$t = \sqrt{\frac{2}{a_T + g}},$$

y el espacio que la Tierra recorrerá en este lapso de tiempo será de

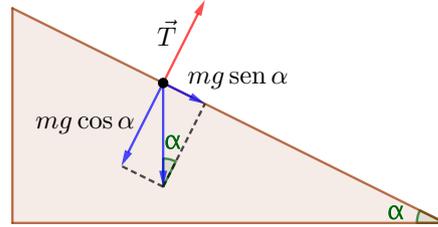
$$r = \frac{1}{2} a_T t^2 = \frac{a_T}{a_T + g} = 1.674 \cdot 10^{-25} \text{ m}.$$

Por lo tanto, en la práctica podemos considerar que la Tierra no se ve afectada por la atracción de la piedra. ■

Ejemplo: Planos inclinados ¿Cómo se mueve un objeto que se desliza sin rozamiento por un plano inclinado?

Si α es la inclinación del plano y el objeto tiene masa m , en cada momento está sometido a dos fuerzas: su peso $\vec{P} = (0, -mg)$ y la reacción del plano que no ofrece ninguna resistencia al avance sobre el propio plano (despreciando el rozamiento), pero sí impide cualquier intento de desplazamiento en la dirección perpendicular al plano. Esto significa que \vec{T} compensa exactamente la componente de \vec{P} perpendicular al plano, cuya intensidad es $mg \cos \alpha$, por

lo que la fuerza resultante es la componente tangencial del peso, y su módulo es $F = mg \operatorname{sen} \alpha$. Por lo tanto, el cuerpo está sometido a una aceleración de módulo $\|\vec{a}\| = g \operatorname{sen} \alpha$ dirigida en la dirección del plano.



Si tomamos un sistema de referencia con un eje sobre el plano inclinado, entonces la ecuación del movimiento es

$$\vec{r} = \vec{r}_0 + \vec{v}_0(t - t_0) - \frac{1}{2}g \operatorname{sen} \alpha (t - t_0)^2.$$

Por ejemplo, si dejamos caer un cuerpo por un plano con inclinación α desde una altura h con velocidad inicial nula, el espacio l que tiene que recorrer para llegar a la base es $l = h / \operatorname{sen} \alpha$, y el tiempo T requerido cumplirá

$$\frac{h}{\operatorname{sen} \alpha} - \frac{1}{2}g \operatorname{sen} \alpha T^2 = 0,$$

luego

$$T = \frac{\sqrt{2h/g}}{\operatorname{sen} \alpha} = \sqrt{\frac{2}{gh}} l. \quad (4.2)$$

Por ejemplo:

¿Cuánto tiempo tardará en llegar al suelo un objeto que se deja caer por un plano inclinado en 60° desde 200 m de altura?

La respuesta es

$$T = \frac{\sqrt{400/9.8}}{1/2} = 12.8 \text{ s.}$$

Observemos que el objeto, en su caída por el plano inclinado, ha avanzado 100 m horizontalmente. Galileo se planteó el problema siguiente:

Si queremos que un objeto recorra una distancia horizontal d , ¿desde qué altura tendríamos que dejarlo caer por una rampa para que tarde en llegar a la base el menor tiempo posible?

Queremos el valor de h que hace mínima la función

$$T(h) = \sqrt{\frac{2}{gh}} \sqrt{d^2 + h^2},$$

o, equivalentemente, su cuadrado:

$$f(h) = T^2(h) = \frac{2}{g} \left(\frac{d^2}{h} + h \right).$$

Para encontrarlo calculamos

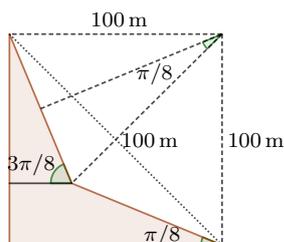
$$f'(h) = \frac{2}{g} \left(-\frac{d^2}{h^2} + 1 \right),$$

y vemos que la derivada es negativa si $h < d$ y positiva si $h > d$, luego f decrece hasta llegar a $h = d$ y a partir de ahí crece, luego el tiempo mínimo se consigue cuando $h = d$, es decir, con una rampa inclinada en 45° . Dicho tiempo mínimo es

$$T = 2\sqrt{d/g}.$$

Por ejemplo, si $d = 100$ m, el tiempo mínimo es $T = 6.39$ s, que mejora sustancialmente los 12.8 s que hemos calculado antes para una la rampa en 60° .

Sin embargo, es posible trasladar 100 m horizontalmente un cuerpo en menos de ese tiempo dejándolo caer por una rampa adecuada. Por ejemplo, basta considerar dos planos inclinados como indica la figura:



En el primer plano inclinado el objeto experimenta una aceleración de módulo $g \sin(3\pi/8)$ y tiene que recorrer un espacio $2 \cdot 100 \sin(\pi/8)$, luego el tiempo t_1 que requerirá para ello cumplirá

$$-200 \sin \frac{\pi}{8} = -\frac{1}{2} \cdot 9.8 \sin \frac{3\pi}{8} t^2,$$

y la solución es $t_1 = 4.112$ s. Cuando llegue a la base del primer plano, la velocidad será

$$v_0 = at_1 = 9.8 \sin \frac{3\pi}{8} 4.112 = 37.228 \text{ m/s}.$$

En el segundo plano inclinado el espacio a recorrer es el mismo y la aceleración es $g \sin(\pi/8)$, pero ahora la velocidad inicial no es nula, sino que es precisamente la velocidad v_0 que acabamos de calcular.⁴ El tiempo t_2 que tardará en recorrerlo cumplirá

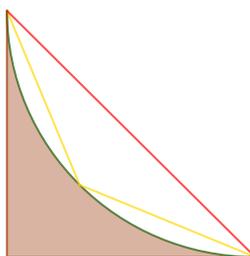
$$-200 \sin \frac{\pi}{8} = -37.228 \cdot t - \frac{1}{2} 9.8 \sin \frac{\pi}{8} t^2 = 0,$$

⁴Aquí hay que suponer que el cambio de plano no supone un "choque" para el objeto que pueda frenarlo. Podemos suponer que la junta está redondeada, de modo que la reacción del plano curva su trayectoria sin modificar el módulo de su velocidad.

y la solución (positiva) de esta ecuación es $t_2 = 1.878$, por lo que el tiempo total que tarda el objeto en llegar a la base es

$$T = t_1 + t_2 = 5.99 \text{ s.}$$

Vemos así que el camino más corto no es el camino más rápido. Aunque con la concatenación de dos planos inclinados el espacio recorrido es mayor, el tiempo empleado en realizar el descenso es cuatro décimas de segundo menor que si la caída se hace en línea recta.



Tal y como señalábamos en la introducción, esto lo observó Galileo en 1638, y vio que considerando más planos inclinados inscritos en una circunferencia el tiempo de descenso se reducía cada vez más, y concluyó que el menor tiempo posible se consigue cuando la trayectoria es un arco de circunferencia.

En la página 334 veremos que el tiempo de caída por el arco de circunferencia es $T = 5.92 \text{ s}$, que, en efecto, es menor que el tiempo de caída por las poligonales inscritas en él, pero también veremos que no es el menor tiempo posible.

Consideremos ahora el otro problema de Galileo que discutimos en la introducción:

Determinar todos los puntos desde los cuales, al dejar caer un objeto por un plano inclinado, éste llega a un punto prefijado en un tiempo prefijado.

Concretamente, si dejamos caer el objeto desde un punto (x, y) , con $y > 0$, por un plano inclinado que termina en $(0, 0)$, el tiempo que tardará en recorrer el plano viene dado por (4.2), que equivale a

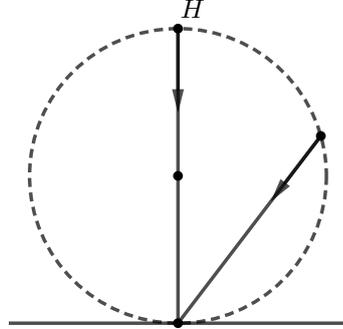
$$T = \frac{\sqrt{2y/g}}{y/\sqrt{x^2 + y^2}},$$

y elevando al cuadrado y operando se convierte en $x^2 + y^2 = Hy$, donde hemos llamado $H = (1/2)gT^2$. Equivalentemente,

$$x^2 + (y - H/2)^2 = (H/2)^2,$$

que es la ecuación de la circunferencia de centro $(0, H/2)$ y radio $H/2$. En conclusión:

Si dejamos caer un objeto desde un punto de una circunferencia hasta su punto más bajo por un plano inclinado, el tiempo que tardará en llegar a dicho punto es el mismo, sea cual sea el punto de partida.



(Se incluye el caso en que el objeto se deja caer verticalmente desde una altura H , en cuyo caso no hace falta ningún plano inclinado.) ■

Ejemplo: La catenaria Tal y como hemos señalado en la introducción, en 1690 Jakob Bernoulli desafió a los matemáticos de su tiempo a encontrar la *catenaria*, es decir, la forma exacta que adopta una cadena cuando se la suspende de sus extremos.

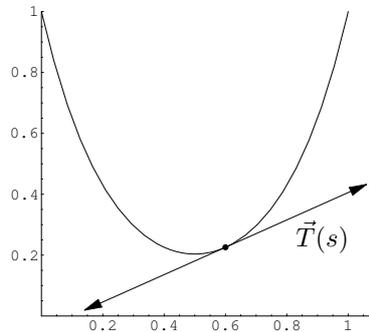
Para resolver el problema tenemos que suponer que la cuerda tiene densidad constante, es decir, que la masa de un segmento de cuerda de longitud s es ρs , para una constante ρ . Llamemos $y(x)$ a la curva que estamos buscando, definida sobre un intervalo $[x_0, x_1]$ con las condiciones $y(x_0) = y_0$, $y(x_1) = y_1$, es decir, suponemos que los extremos de la cuerda son dos puntos (x_0, y_0) , (x_1, y_1) . Además exigimos que la longitud de la cuerda tenga un valor prefijado L , es decir, que se tiene que cumplir

$$\int_{x_0}^{x_1} \sqrt{1 + y'(x)^2} dx = L.$$

Consideramos la parametrización por el arco $\phi(s) = (x(s), y(s))$ y llamamos además $\alpha(s)$ al ángulo que forma con la horizontal la tangente a la cuerda en $\phi(s)$. La interpretación geométrica de la derivada equivale a que

$$y'(x) = \tan \alpha(s(x)).$$

Consideremos un punto cualquiera de la cuerda (es decir, un punto $\phi(s)$, para un cierto valor de $s \in [0, L]$). La fuerza de cohesión que mantiene unida la cuerda por dicho punto —la fuerza que desaparecería si cortáramos la cuerda— está representada por dos vectores de tensión opuestos, $\vec{T}(s)$ y $-\vec{T}(s)$, donde convenimos que $\vec{T}(s)$ es el que tiene su componente horizontal positiva.



Concretamente, $\vec{T}(s)$ es la fuerza que la parte de la cuerda posterior a s ejerce sobre la parte anterior y $-\vec{T}(s)$ es la fuerza que la parte anterior de la cuerda ejerce sobre la parte posterior. Si, como en la figura, consideramos un punto donde la curva es creciente, podemos decir, más descriptivamente, que $\vec{T}(s)$ es la fuerza con la que la parte superior de la cuerda “sostiene” de la parte inferior, mientras que $-\vec{T}(s)$ es la fuerza con la que la parte inferior de la cuerda “estira” de la parte superior. En los puntos donde la curva es decreciente se invierten los papeles. Que ambas tensiones tengan la misma dirección e intensidad, pero sentidos opuestos, es consecuencia de la tercera ley de Newton: la fuerza que ejerce una parte de la cuerda sobre la otra ha de ser exactamente la opuesta de la que la otra ejerce sobre la primera. Es claro que su dirección debe ser la de la tangente a la curva⁵, luego podemos descomponer $\vec{T}(s)$ como

$$\vec{T}(s) = (T(s) \cos \alpha(s), T(s) \sin \alpha(s)).$$

Alguien podría pensar que sobre el punto $\alpha(s)$ se ejerce una tercera fuerza, a saber, el peso, pero debemos recordar que los puntos no tienen masa, por lo que tampoco tienen peso. Para tener en cuenta el peso hemos de considerar, no un punto, sino un arco de cuerda, digamos el correspondiente al intervalo de parámetros $[s, s + \Delta s]$. Dicho arco está sometido a tres fuerzas: su peso

$$\vec{P} = (0, -\rho g \Delta s)$$

y las tensiones con las que la cuerda “tira” de sus extremos, que son $\vec{T}(s + \Delta s)$, $-\vec{T}(s)$. La condición de equilibrio es que la suma total sea nula:⁶

$$\vec{T}(s + \Delta s) - \vec{T}(s) + \vec{P} = \vec{0}.$$

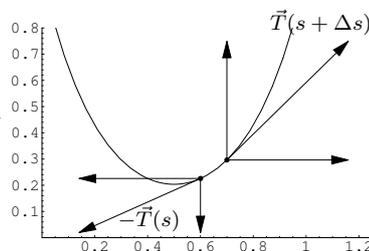
Más explícitamente:

$$T(s + \Delta s) \cos \alpha(s + \Delta s) - T(s) \cos \alpha(s) = 0,$$

$$T(s + \Delta s) \sin \alpha(s + \Delta s) - T(s) \sin \alpha(s) = \rho g \Delta s.$$

Si hacemos tender $\Delta s \rightarrow 0$ vemos que

$$\frac{d(T \cos \alpha)}{ds} = 0, \quad \frac{d(T \sin \alpha)}{ds} = \rho g.$$



Esto es cierto para todo punto $s \in]0, L[$. La primera condición equivale a que la componente horizontal de la tensión es constante:

$$T(s) \cos \alpha(s) = \kappa > 0.$$

⁵La tensión de una cuerda recta se ejerce en la dirección determinada por ella misma, y una curva diferenciable, vista al microscopio en un entorno de uno de sus puntos, se confunde con su recta tangente, lo que muestra que ésta ha de ser la dirección de las tensiones en ese punto.

⁶Observemos que las tensiones se aplican en puntos distintos y que el peso no se aplica en ningún punto en particular. La condición de suma nula sería obviamente necesaria si el arco fuera rígido, pero con mayor razón ha de ser necesaria si es flexible, ya que una suma de fuerzas que movería un arco rígido, con mayor razón movería a un arco flexible.

Sustituyendo $T(s)$ en la segunda igualdad obtenemos que

$$\frac{d \tan \alpha(s)}{ds} = \frac{\rho g}{\kappa} = \mu.$$

Equivalentemente:

$$\mu = \frac{d \tan \alpha(s(x))}{dx} \frac{dx}{ds} = \frac{dy'(x)}{dx} \frac{1}{\sqrt{1 + y'(x)^2}}.$$

En resumen, hemos llegado a la ecuación diferencial:

$$\frac{y''(x)}{\sqrt{1 + y'(x)^2}} = \mu.$$

Considerándola como ecuación en $u = y'$ y separando variables obtenemos

$$\operatorname{arsenh} y'(x) = \mu x + c_1,$$

es decir,

$$y'(x) = \operatorname{senh}(\mu x + c_1),$$

e integrando a su vez esta ecuación llegamos a

$$y(x) = \frac{\cosh(\mu x + c_1)}{\mu} + c_2.$$

Ésta es la ecuación de la catenaria, que, como vemos, es esencialmente el coseno hiperbólico, modificado con parámetros μ , c_1 y c_2 que nos permiten ajustar la longitud y los extremos. Concretamente, la condición sobre la longitud nos da que

$$L = \int_{x_0}^{x_1} \sqrt{1 + y'(x)^2} dx = \int_{x_0}^{x_1} \sqrt{1 + \operatorname{senh}^2(\mu x + c_1)} dx =$$

$$\int_{x_0}^{x_1} \cosh(\mu x + c_1) dx = \left[\frac{\operatorname{senh}(\mu x + c_1)}{\mu} \right]_{x_0}^{x_1} = \frac{\operatorname{senh}(\mu x_1 + c_1)}{\mu} - \frac{\operatorname{senh}(\mu x_0 + c_1)}{\mu},$$

luego, en total, tenemos tres ecuaciones que determinan las tres constantes:

$$\frac{\cosh(\mu x_0 + c_1)}{\mu} + c_2 = y_0, \quad \frac{\cosh(\mu x_1 + c_1)}{\mu} + c_2 = y_1,$$

$$\frac{\operatorname{senh}(\mu x_1 + c_1) - \operatorname{senh}(\mu x_0 + c_1)}{\mu} = L.$$

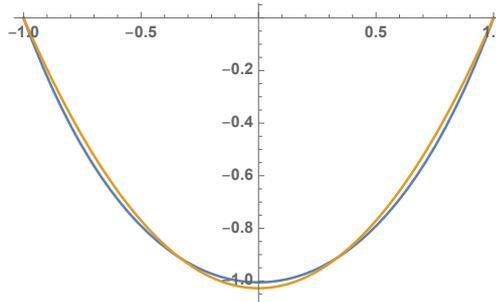
En general, calcular los parámetros de la catenaria a partir de las condiciones iniciales requiere resolver numéricamente este sistema de ecuaciones. Consideremos, por ejemplo, las condiciones $x(-1) = x(1) = 0$ y $L = 3$. Es claro entonces que la catenaria tiene que ser simétrica respecto del eje vertical, lo cual se traduce en que $c_1 = 0$.

La condición sobre la longitud es

$$\sinh(\mu) - \sinh(-\mu) = 3\mu,$$

o también $\sinh(\mu) = 3\mu/2$, cuya solución es $\mu = 1.62213\dots$, de donde

$$c_1 = c_2 = -\frac{\cosh \mu}{\mu} = 1.621739\dots$$



La figura muestra la catenaria y la parábola de longitud 3 que pasan por los puntos $(\pm 1, 0)$ (la parábola es la que cae más bajo). En la introducción hemos mostrado las gráficas de catenarias y parábolas de otras longitudes. ■

Ejemplo: La ley de Hooke Los muelles proporcionan una técnica para medir fuerzas gracias a que verifican la llamada ley de Hooke. Un muelle tiene la propiedad de que se puede estirar o contraer bajo la acción de una fuerza, pero es capaz de regresar a su posición inicial una vez ha cesado la fuerza causante de la deformación. Por simplificar la situación, supongamos que metemos un muelle en un tubo, de modo que no pueda realizar ninguna clase de desplazamiento más que en sentido longitudinal, y fijamos uno de sus extremos dejando libre el otro. Tomamos un sistema de referencia de modo que el muelle esté sobre el eje x y su extremo libre ocupe la posición $x = 0$ cuando no actúa ninguna fuerza.



La *ley de Hooke* afirma que el muelle ejerce sobre su extremo libre una fuerza proporcional a la distancia en que ha sido desplazado de la posición que ocuparía en ausencia de fuerzas. Como esta fuerza está dirigida en sentido contrario al desplazamiento, respecto al sistema de referencia que hemos fijado se expresará como $F = -kx$, donde k es una constante que depende del muelle. Así, por ejemplo, para mantener el muelle en la posición de la figura de la derecha, es necesario mantener aplicada en su extremo una fuerza $F' = kx$ (hacia la izquierda) que compense la fuerza hacia la derecha $-kx$ que ejerce el muelle.

La constante k puede determinarse sin más que aplicar al muelle una fuerza de intensidad conocida (por ejemplo, colgando de él un cuerpo de masa conocida). Una vez determinada, el muelle se convierte en un *dinamómetro*, es decir, en un aparato de medida de fuerzas. Para medir una fuerza se la hace actuar sobre el extremo del muelle, se mide el desplazamiento que produce y se multiplica por la constante k . En particular, un muelle puede usarse como báscula, ya sea poniéndolo debajo de un plato para que cualquier objeto depositado en él oprima el muelle con la fuerza de su peso, ya sea colgando de él un plato para que el peso lo estire.

En realidad la ley de Hooke sólo es válida para desplazamientos pequeños. Si son demasiado grandes, la fuerza del muelle puede ser menor de la prevista por la ley de Hooke, y si son excesivamente grandes el muelle se deformará de forma permanente o incluso se podría romper.

Supongamos ahora que sujetamos un objeto de masa m en el extremo libre de un muelle, lo alejamos de su posición de equilibrio hasta una posición x_0 y lo soltamos. La segunda ley de Newton nos da entonces que $mx'' = -kx$, o también

$$x'' = -\frac{k}{m}x, \quad x(0) = x_0, \quad x'(0) = 0.$$

Esto es una ecuación diferencial de las estudiadas en el teorema 1.35, luego su solución es

$$x = x_0 \cos(\sqrt{k/m} t).$$

En efecto, ésta es la solución de la familia de soluciones consideradas que cumple las dos condiciones iniciales. Vemos así que el objeto situado en el extremo del muelle oscilará para siempre alrededor del punto de equilibrio, con periodo

$$T = 2\pi \sqrt{\frac{m}{k}}$$

(el tiempo que tarda en ir y venir) y frecuencia

$$\nu = \frac{1}{T} = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$$

(el número de oscilaciones por segundo). En la práctica las oscilaciones se atenuarán por el rozamiento hasta que el muelle se estabilice en su posición de equilibrio.

La situación anterior supone que el muelle está oscilando horizontalmente, de modo que el peso del cuerpo situado en su extremo no afecta al movimiento, pues es compensado por la reacción de la superficie sobre la que se apoya. Supongamos ahora que el muelle está en vertical, con el cuerpo colgando de él. Si el objeto está en reposo, su posición no será $x = 0$, sino $x = l < 0$ (entendiendo como es habitual que las posiciones negativas son las situadas por debajo de la posición de $x = 0$, que corresponde a la posición de equilibrio del muelle) de modo que $-kl - mg = 0$.

Si desplazamos el cuerpo hasta una posición $x_0 \neq l$ y lo soltamos, ahora en la segunda ley de Newton tenemos que tener en cuenta que la fuerza resultante que actúa sobre el cuerpo es su peso y la ejercida por el muelle:

$$-mg - kx = ma,$$

y esto puede expresarse en la forma

$$-k(x - l) = mx'' = m(l + x)'', \quad x(0) - l = x_0 - l, \quad x'(0) = 0.$$

Por lo tanto, la función $x - l$ satisface también una ecuación diferencial del tipo considerado en el teorema 1.35, cuya solución es

$$x = l + (x_0 - l) \cos(\sqrt{k/m}t).$$

Vemos que el muelle oscila con el mismo periodo que cuando no influía el peso, pero alrededor de $x = l$, en lugar de alrededor de $x = 0$. ■

El péndulo simple Según hemos contado en la introducción, Galileo observó que el periodo de oscilación de un péndulo es independiente de la amplitud de sus oscilaciones, si bien esto sólo es cierto para amplitudes pequeñas. Vamos a ver por qué. Para ello consideremos un objeto de masa m colgado de una cuerda de longitud l . Lo desplazamos un poco desde la vertical con la cuerda tensa y lo dejamos caer. Observemos que en cada instante actúan dos fuerzas sobre el péndulo: su peso \vec{P} y la tensión de la cuerda \vec{T} . En efecto, la cuerda ejerce una fuerza, pues si el péndulo se moviera únicamente bajo la acción de su peso, simplemente caería. Además, es obvio que una cuerda no puede hacer más que tirar en su propia dirección, es decir, que \vec{T} tiene la dirección de la cuerda y apunta hacia su extremo fijo.

El peso es $\vec{P} = (0, -mg)$, pero podemos descomponerlo en una componente radial \vec{P}_r y una componente transversal \vec{P}_t . Si la posición del objeto es $\vec{r} = (l \sin \theta, -l \cos \theta)$, entonces la componente radial tiene módulo $P_r = mg \cos \theta$ y la componente transversal $P_t = mg |\sin \theta|$. Por otra parte:

$$\vec{v} = (l\theta' \cos \theta, l\theta' \sin \theta),$$

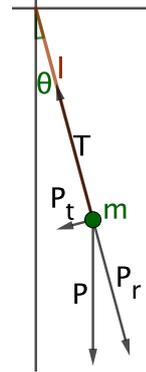
$$\vec{a} = (l\theta'' \cos \theta, l\theta'' \sin \theta) - (l\theta'^2 \sin \theta, -l\theta'^2 \cos \theta) = l\theta'' \vec{\tau} - \theta'^2 \vec{r},$$

donde $\vec{\tau} = (\cos \theta, \sin \theta)$ es el vector tangente unitario. La segunda ley de Newton nos da la ecuación

$$\vec{T} + \vec{P} = m\vec{a},$$

que podemos descomponer en sus componentes radial y transversal:

$$\vec{T} + \vec{P}_r = -m\theta'^2 \vec{r}, \quad \vec{P}_t = -ml\theta'' \vec{\tau}.$$



La primera ecuación expresa que la suma de la tensión de la cuerda más la componente radial del peso debe ser igual a la masa del péndulo por la aceleración normal del péndulo. Expresando

$$\vec{T} = -\frac{T}{l}\vec{r}, \quad \vec{P}_r = \frac{mg}{l}\cos\theta\vec{r},$$

se reduce a que

$$-\frac{T}{l} + \frac{mg}{l}\cos\theta = -m\theta'^2,$$

de donde el módulo de la tensión de la cuerda es $T = mg\cos\theta + m\theta'^2l$. Pero la ecuación más interesante es la segunda. Observemos que $\vec{P}_t = -mg\sin\theta\vec{\tau}$, pues si $\theta > 0$ (como en la figura), el vector $\vec{\tau}$ tiene sus coordenadas positivas, mientras que \vec{P}_t las tiene negativas, y viceversa. Por lo tanto la segunda ecuación nos da:

$$\theta'' = -\frac{g}{l}\sin\theta. \quad (4.3)$$

Tenemos así una ecuación diferencial de segundo orden que determina la función $\theta(t)$ y, por consiguiente, la posición del péndulo en cada instante, si le añadimos las condiciones iniciales $\theta(0) = \theta_0$, $\theta'(0) = 0$.

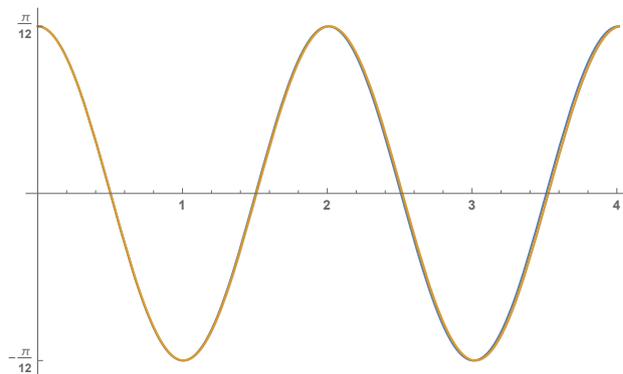
De momento no sabemos resolver esta ecuación diferencial (lo haremos en la sección 6.5), pero si la amplitud máxima θ_0 es pequeña, la aproximación $\sin\theta \approx \theta$ es razonable, y la ecuación diferencial se convierte en

$$\theta'' = -\frac{g}{l}\theta,$$

que tampoco es trivial, pero que hemos resuelto en el teorema 1.35. Su solución es

$$\theta = \theta_0 \cos\left(\sqrt{\frac{g}{l}}t\right),$$

pues, ciertamente, con esta elección de las constantes, se cumplen las condiciones iniciales $\theta(0) = \theta_0$ y $\theta'(0) = 0$. La figura siguiente muestra la gráfica de la solución $\theta(t)$ de la ecuación diferencial exacta (con el término $\sin\theta$) y la solución aproximada que hemos encontrado para $l = 1$ m y $\theta_0 = 15^\circ$ (que es precisamente el ángulo que muestra la figura anterior):



Vemos que son prácticamente indistinguibles. Según la solución aproximada que hemos encontrado, el periodo de oscilación es

$$T = 2\pi\sqrt{\frac{l}{g}},$$

lo cual confirma los experimentos de Galileo: es independiente de la amplitud θ_0 y proporcional a \sqrt{l} . Éste es el fundamento de los relojes de péndulo y de los metrónomos.

Notemos por último que tanto la solución exacta como la aproximada predicen que el péndulo continuará oscilando indefinidamente siempre con la misma amplitud. Esto no es cierto en la práctica, sino que un péndulo va disminuyendo poco a poco su amplitud (pero no su periodo) hasta que se para, pero esto se debe a que sobre él actúa una tercera fuerza: la fuerza de rozamiento del aire que va frenándolo paulatinamente y que no hemos tenido en cuenta en nuestro planteamiento. ■

4.3 La ley de gravitación universal

En esta sección probaremos que la ley de gravitación universal de Newton implica las leyes de Kepler sobre el movimiento planetario. Recordemos su enunciado:

Ley de gravitación universal *Dos cuerpos cualesquiera se atraen con una fuerza directamente proporcional a sus masas e inversamente proporcional al cuadrado de la distancia que los separa.*

Consideramos un sistema de referencia con el Sol en el origen de coordenadas y llamamos \vec{r} al vector de posición de un planeta o, más en general, de cualquier cuerpo que se mueva sometido a la gravedad solar. Si llamamos M a la masa del Sol y m a la del planeta, la ley de gravitación universal afirma que éste experimenta una fuerza dirigida hacia el origen de coordenadas dada por

$$\vec{F} = -\frac{GMm}{r^3}\vec{r},$$

para cierta constante G . Notemos que así el módulo de \vec{F} es $F = GMm/r^2$, como requiere la ley de gravitación. Suponemos que ninguna otra fuerza actúa sobre el planeta. La segunda ley de Newton nos da la ecuación diferencial

$$-\frac{GMm}{r^3}\vec{r} = m\vec{a}.$$

No es fácil resolver esta ecuación directamente, pero podemos extraer mucha información de ella sobre cómo tienen que ser sus soluciones. Para ello introducimos el *momento angular* de un objeto respecto de un sistema de referencia dado, que se define como

$$\vec{L} = \vec{r} \times m\vec{v}.$$

Su interés radica en que (cuando la fuerza se dirige hacia el origen de coordenadas) se mantiene constante, pues

$$\frac{d\vec{L}}{dt} = \vec{v} \times m\vec{v} + \vec{r} \times m\vec{a} = \vec{r} \times \vec{F} = \vec{0},$$

pues \vec{F} y \vec{r} son proporcionales.

Como \vec{L} es ortogonal a \vec{r} , concluimos que el planeta se va a mover siempre en el plano que pasa por el origen de coordenadas y es ortogonal a \vec{L} . Podemos elegir el sistema de referencia de modo que dicho plano sea el plano XY, con lo que podemos expresar \vec{r} en coordenadas polares:

$$\vec{r} = (\rho \cos \theta, \rho \sin \theta, 0).$$

A su vez, llamando $\omega = \theta'$,

$$\vec{v} = \rho'(\cos \theta, \sin \theta, 0) + \rho\omega(-\sin \theta, \cos \theta, 0), \quad (4.4)$$

luego

$$\begin{aligned} \vec{L} &= \rho(\cos \theta, \sin \theta, 0) \times (m\rho'(\cos \theta, \sin \theta, 0) + m\rho\omega(-\sin \theta, \cos \theta, 0)) \\ &= m\rho^2\omega(\cos \theta, \sin \theta, 0) \times (-\sin \theta, \cos \theta, 0) = (0, 0, m\rho^2\omega). \end{aligned}$$

Por consiguiente,

$$L = m\rho^2\omega$$

tiene que ser invariante en el tiempo. Volvemos a derivar llamando $\alpha = \omega' = \theta''$:

$$\vec{a} = (\rho'' - \rho\omega^2)(\cos \theta, \sin \theta, 0) + (2\rho'\omega + \rho\alpha)(-\sin \theta, \cos \theta, 0).$$

La fuerza gravitatoria es:

$$\vec{F} = -\frac{GMm}{\rho^2}(\cos \theta, \sin \theta, 0).$$

Teniendo en cuenta que los vectores $(\cos \theta, \sin \theta, 0)$ y $(-\sin \theta, \cos \theta, 0)$ son ortogonales, la igualdad $\vec{F} = m\vec{a}$ equivale a las dos ecuaciones diferenciales:

$$\rho'' - \rho\omega^2 = -\frac{GM}{\rho^2}, \quad 2\rho'\omega + \rho\alpha = 0. \quad (4.5)$$

La solución es complicada, pero de momento vamos a tratar de determinar la forma de la trayectoria, aunque sea con otra parametrización. Por ello, en lugar de calcular las funciones $\rho(t)$ y $\theta(t)$ calcularemos la parametrización $\rho(\theta)$.

Ahora bien, hay un caso en el que esta parametrización es imposible: Si $\omega_0 = 0$, por (4.4) tenemos que \vec{v}_0 es nula o proporcional a \vec{r}_0 . El hecho de que L sea constante implica que si $\omega_0 = 0$, de hecho $\omega = 0$ en todo instante t , luego también $\alpha = 0$ y la segunda ecuación de (4.5) se cumple trivialmente. Así pues,

en este caso la solución de las ecuaciones es una recta de la forma $(\rho(t), \theta_0)$, donde ρ está determinada por la ecuación

$$\rho'' = -\frac{GM}{\rho^2} \quad (4.6)$$

con las condiciones iniciales $\rho_0 = \|\vec{r}_0\|$ y $\rho'_0 = \pm\|\vec{v}_0\|$ (por (4.4)).

Notemos que, según hemos observado, la invarianza de L implica que si $\omega(t) = 0$ en un instante t , entonces $\omega = 0$ en todo instante, luego, recíprocamente, si $\omega_0 \neq 0$ entonces ω no se anula nunca. Esto nos permite usar la función $\theta(t)$ como cambio de parámetro, que nos da la reparametrización $\rho(\theta)$. Claramente

$$\rho' = \rho'_\theta \omega \quad \text{y} \quad \rho'' = \rho''_\theta \omega^2 + \rho'_\theta \alpha.$$

Sustituimos estas igualdades en (4.5) y eliminamos α en la primera usando la segunda. El resultado es la ecuación

$$\left(\rho'' - \frac{2\rho'^2}{\rho} - \rho \right) \omega^2 = -\frac{GM}{\rho^2},$$

donde ahora todas las derivadas son respecto de θ y no respecto de t . No obstante, la presencia de ω nos obliga a considerar ambos miembros como funciones de t (pues la expresión entre paréntesis es una función de θ compuesta con la función $\theta(t)$).

Sin embargo, al multiplicar ambos miembros por $m^2\rho^4$ queda

$$\frac{\rho''}{\rho^2} - \frac{2\rho'^2}{\rho^3} - \frac{1}{\rho} = -\frac{GMm^2}{L^2}, \quad (4.7)$$

donde L una constante, luego todo el segundo miembro es constante y la igualdad sigue siendo válida si consideramos el primer miembro como función de θ .

Pero el teorema 1.37 nos da entonces que $\rho(\theta)$ es de la forma

$$\rho = \frac{L^2}{GMm^2} \frac{1}{1 + \epsilon \cos(\theta + \phi)},$$

para ciertas constantes $\epsilon \geq 0$ y ϕ , que determina una cónica de excentricidad ϵ con uno de sus focos en el origen de coordenadas, es decir, en el Sol. Con esto hemos probado la primera ley de Kepler generalizada:

Los cuerpos sometidos únicamente a la tracción gravitatoria de otro cuerpo siguen trayectorias rectas o secciones cónicas, de modo que cualquier sección cónica puede ser la trayectoria de un cuerpo, si la posición y la velocidad de éste en un instante dado son las adecuadas.

Si elegimos el sistema de referencia de modo que el perihelio corresponda al ángulo $\theta = 0$, entonces $\phi = 0$. Si v_0 es la velocidad del planeta el en perihelio y ρ_0 su distancia al Sol, entonces

$$\rho_0 = \frac{L^2}{GMm^2} (1 + \epsilon)^{-1} = \frac{m^2 \rho_0^2 v_0^2}{GMm^2} (1 + \epsilon)^{-1}, \quad (4.8)$$

luego la excentricidad de la órbita en función de las condiciones iniciales en el perihelio es

$$\epsilon = \frac{\rho_0 v_0^2}{GM} - 1.$$

Para órbitas elípticas, el *afelio* o punto de mayor distancia al Sol se alcanza cuando $\theta = \pi$, luego

$$\rho_1 = \frac{L^2}{GMm^2} (1 - \epsilon)^{-1} = \rho_0 \frac{1 + \epsilon}{1 - \epsilon}.$$

El semieje mayor de la órbita es

$$a = \frac{\rho_0 + \rho_1}{2} = \frac{\rho_0}{1 - \epsilon} = \frac{L^2}{GMm^2} \frac{1}{(1 + \epsilon)^2}. \quad \blacksquare$$

Ahora es fácil demostrar la segunda ley de Kepler, es decir, que el radio que une el cuerpo que orbita con el Sol barre áreas iguales en tiempos iguales. En efecto, según el teorema 2.6, el área de una cónica comprendida entre dos ángulos θ_0 y θ_1 es

$$A = \frac{1}{2} \int_{\theta_0}^{\theta_1} \rho^2 d\theta.$$

Hacemos el cambio $\theta = \theta(t)$, donde t es el tiempo, y el resultado es

$$\frac{1}{2} \int_{t_0}^{t_1} \rho^2(\theta(t)) \theta'(t) dt = \frac{1}{2} \int_{t_0}^{t_1} \rho^2(t) \omega(t) dt = \frac{1}{2m} \int_{t_0}^{t_1} L dt = \frac{L}{2m} (t_1 - t_0).$$

Así pues, el área barrida es proporcional al tiempo empleado en barrerla. Por lo tanto, para saber en qué punto de la órbita estará un planeta en un momento dado (conociendo su posición en un instante inicial) puede calcularse conociendo su *velocidad areolar*, es decir, el área barrida por unidad de tiempo, que según el cálculo precedente es

$$v_a = \frac{L}{2m}.$$

En el caso de una órbita elíptica puede calcularse a partir del periodo de rotación T , es decir, del tiempo que tarda el cuerpo en dar una vuelta completa, pues entonces $v_a = A/T$, donde A es el área de la elipse, que, según hemos visto en la página 86, viene dada por $A = \pi ab$, donde a y b son los semiejes. Hemos visto que

$$a = \frac{L^2}{GMm^2} \frac{1}{1 - \epsilon^2}.$$

En cuanto al semieje menor, según se ve en la sección 7.1 de [IGE], como ya hemos tenido en cuenta en la sección 2.3, es $b = a\sqrt{1 - \epsilon^2}$. El área de la elipse es, por consiguiente,

$$A = \pi ab = \pi a^2 \sqrt{1 - \epsilon^2} = \frac{\pi}{G^2 M^2} \frac{L^4}{m^4 (1 - \epsilon^2)^2} \sqrt{1 - \epsilon^2}.$$

Según hemos calculado, la velocidad areolar es $A/T = L/(2m)$, luego

$$T = \frac{2\pi}{G^2 M^2} \frac{L^3 \sqrt{1 - \epsilon^2}}{m^3 (1 - \epsilon^2)^2}.$$

Combinando las expresiones de a y T obtenemos

$$\frac{T^2}{a^3} = \frac{4\pi^2}{GM}, \quad (4.9)$$

que es la tercera ley de Kepler: los cuadrados de los periodos de revolución son proporcionales a los cubos de los semiejes mayores, con el añadido de que ahora tenemos una expresión sencilla para la constante de proporcionalidad. ■

El experimento de Cavendish Observemos ahora que la ley de gravitación universal que —según acabamos de comprobar— explica el movimiento de los cuerpos celestes, también explica el comportamiento familiar de la gravedad terrestre. Un cuerpo situado cerca de la superficie de la Tierra es atraído por ésta con una fuerza dirigida hacia su centro y de módulo

$$F = \frac{GM_T m}{r^2},$$

donde M_T la masa de la Tierra t r es la distancia del cuerpo al centro de la Tierra. Expresándola como $r = R_T + h$, donde R_T es el radio de la Tierra,

$$F = \frac{GM_T m}{R_T^2} \left(\frac{R_T}{R_T + h} \right)^2 = mg \left(\frac{R_T}{R_T + h} \right)^2,$$

donde

$$g = \frac{GM_T}{R_T^2} = 9.80665 \text{ N/m} \quad (4.10)$$

es la intensidad del campo gravitatorio terrestre que ya hemos manejado en varias ocasiones. Ahora bien, cuando la altura h es pequeña en proporción con el radio terrestre R_T , el segundo factor es prácticamente 1 y se puede despreciar. Por ejemplo, para una altura de 5 km sobre el nivel del mar tenemos que

$$\left(\frac{R_T}{R_T + h} \right)^2 = \left(\frac{6378.1}{6383.1} \right)^2 = 0.998434,$$

lo que cambia el valor de g de 9.8 N/m a 9.79 N/m. Por ello, para movimientos de cuerpos que no se eleven excesivamente, es razonable suponer que g es constante, como hemos hecho en la sección anterior.

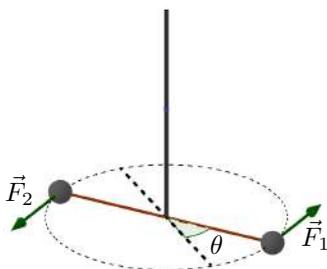
Pero ahora hemos encontrado la relación $GM = gR_T^2$, donde el miembro derecho es conocido, por lo que determinar la constante de gravitación universal G es equivalente a calcular la masa M de la Tierra.

Así, si Eratóstenes se hizo famoso por haber calculado con bastante aproximación el radio (y, por consiguiente, el volumen) de la Tierra sin salir de

Alejandría, en 1798 Henry Cavendish hizo algo mucho más difícil sin salir de su casa: pesó la Tierra. Más precisamente, Cavendish calculó la densidad de la Tierra, lo cual, conocido su radio, es equivalente a calcular su masa, y, por la relación precedente, es equivalente a calcular la constante G . En realidad Cavendish no consideró para nada la constante G . En su época no se consideraba relevante y a menudo ni siquiera se escribía, sino que se hablaba meramente de proporcionalidad entre la fuerza, las masas y las distancias. Pero sus mediciones permiten calcular G exactamente igual que M (de hecho, con menos procesamiento matemático de los datos experimentales).

Cavendish se valió de una balanza de torsión, un dispositivo inventado en 1777 por el físico francés Charles-Augustin de Coulomb para medir la intensidad de la fuerza eléctrica, pero al parecer el reverendo británico John Michell la descubrió independientemente con el propósito de medir la intensidad de la gravedad y construyó una, pero murió antes de poder emplearla, y Cavendish la heredó.

Una balanza de torsión no es más que una varilla rígida de longitud L colgada de un cable por su centro con dos bolas iguales de masa m en sus extremos, de modo que la masa de la barra sea despreciable frente a m .



Consideramos un sistema de referencia con origen en el punto donde el cable sujeta la barra y con el eje Z siguiendo la dirección del cable. Usar el dispositivo como balanza requiere que sobre las bolas actúen únicamente las fuerzas siguientes:

- Los pesos (iguales) $\vec{P} = (0, 0, -mg)$.
- Las fuerzas \vec{T}_1 y \vec{T}_2 que la varilla ejerce sobre ellas.
- Otras fuerzas $\vec{F}_1 = -\vec{F}_2$ que vamos a suponer que no tienen componente vertical. (Son las fuerzas que mediremos con la balanza.)

El hecho de que los pesos de las bolas sean iguales, junto con que las fuerzas \vec{F}_i no tengan componente vertical, se traduce claramente en que las bolas se mantendrán en equilibrio como en una balanza, sin subir ni bajar, lo que significa que las componentes verticales de las fuerzas \vec{T}_i serán ambas iguales a $-\vec{P}$, es decir, que uno de los efectos de la varilla sobre las bolas es el de compensar su peso. Más aún, el hecho de que las fuerzas \vec{F}_i tengan sentidos opuestos implica que sus componentes radiales también tendrán sentidos opuestos, y es

fácil convencerse entonces de que el centro de la varilla permanecerá inmóvil, el cable se mantendrá siempre en posición vertical y el único movimiento posible de las bolas será girar en la circunferencia de centro el origen de coordenadas y radio $L/2$.

Así pues, los vectores de posición de las bolas serán de la forma

$$\vec{r}_1 = \frac{L}{2}(\cos \theta_1, \sin \theta_1, 0), \quad \vec{r}_2 = \frac{L}{2}(\cos \theta_2, \sin \theta_2, 0),$$

donde $\theta_2 = \theta_1 + \pi$. Las velocidades serán:

$$\vec{v}_i = \frac{L}{2}\omega(-\sin \theta_i, \cos \theta_i, 0)$$

donde $\omega = \theta'_i$ (que no depende de i), y las aceleraciones:

$$\vec{a}_i = \frac{L}{2}\alpha(-\sin \theta_i, \cos \theta_i, 0) - \frac{1}{2}\omega^2(\cos \theta_i, \sin \theta_i, 0),$$

donde $\alpha = \omega' = \theta''_i$. La segunda ley de Newton afirma que la fuerza total \vec{f}_i que actúa sobre cada bola cumple:

$$\vec{f}_i = \vec{F}_i + \vec{T}_i + \vec{P} = m\vec{a}_i.$$

En realidad, el movimiento de la balanza está determinado por la evolución de los ángulos θ_i , y las fuerzas contienen información irrelevante (los pesos no contribuyen y las fuerzas radiales son simplemente las necesarias para que las bolas giren, pero no afectan a los ángulos θ_i). Por ello el análisis del movimiento se simplifica si en lugar de considerar fuerzas consideramos los momentos que determinan, donde el *momento* de una fuerza \vec{f} que actúa sobre un cuerpo situado en la posición \vec{r} (respecto del sistema de referencia considerado) se define como $\vec{M} = \vec{r} \times \vec{f}$. En nuestro caso:

$$\vec{M}_i = \vec{r}_i \times \vec{f}_i = \frac{L}{2}(\cos \theta_i, \sin \theta_i, 0) \times m\frac{L}{2}\alpha(-\sin \theta_i, \cos \theta_i, 0),$$

luego

$$\vec{M}_i = (0, 0, \frac{1}{4}mL^2\alpha).$$

Pero es más práctico trabajar con la suma de ambos momentos:

$$\vec{M} = \vec{M}_1 + \vec{M}_2 = (0, 0, \frac{1}{2}mL^2\alpha).$$

Este momento total será la suma de los momentos totales correspondientes a los tres pares de fuerza. El de los pesos se cancela:

$$\vec{r}_1 \times (0, 0, -mg) + \vec{r}_2 \times (0, 0, -mg) = \vec{0},$$

pues $\vec{r}_2 = -\vec{r}_1$. Para calcular el de las fuerzas \vec{F}_i las descomponemos en sus componentes tangencial y radial:

$$\vec{F}_i = F_t^i(-\sin \theta_i, \cos \theta_i, 0) + F_r^i(\cos \theta_i, \sin \theta_i, 0),$$

y observamos que, como $\vec{F}_2 = -\vec{F}_1$ y $\vec{r}_2 = -\vec{r}_1$, ahora los momentos correspondientes a las fuerzas \vec{F}_i no se cancelan, sino que son iguales:

$$\vec{M}_F = 2\vec{M}_{F_1} = 2\frac{L}{2}(\cos\theta_1, \sin\theta_1, 0) \times F_t^1(-\sin\theta_1, \cos\theta_1, 0) = (0, 0, LF_t^1).$$

De hecho, es fácil ver que $F_t^1 = F_t^2$, por lo que podemos referirnos a este valor común como F_t , que es el módulo de la componente tangencial de cualquiera de las fuerzas \vec{F}_i , con signo positivo si ambas apuntan en sentido antihorario y negativo en caso contrario.

Similarmente descomponemos

$$\vec{T}_i = T_t^i(-\sin\theta_i, \cos\theta_i, 0) + T_r^i(\cos\theta_i, \sin\theta_i, 0) + (0, 0, mg),$$

donde hemos usado que la componente vertical tiene que ser exactamente la necesaria para compensar el peso. Lo único que nos interesa señalar aquí es que los momentos correspondientes a las componentes verticales

$$\vec{r}_1 \times (0, 0, mg) + \vec{r}_2 \times (0, 0, mg) = \vec{0}$$

se cancelan, mientras que los de las componentes radiales se anulan, luego el momento total de las fuerzas \vec{T}_i es de la forma $(0, 0, M_T)$. Como todos los momentos son verticales podemos trabajar únicamente con sus terceras componentes, y así, el resumen de lo que hemos obtenido es:

$$LF_t + M_T = \frac{1}{2}mL^2\alpha,$$

donde F_t es el módulo de la componente tangencial de cualquiera de las fuerzas \vec{F}_i (con el signo correspondiente al sentido de giro) y $(0, 0, M_T)$ es el momento total correspondiente a las fuerzas \vec{T}_i .

La balanza de torsión nos permitirá medir F_T si sabemos calcular M_T , y éste es el motivo por el que tienen interés las balanzas de torsión, y es que M_T es fácil de calcular. En principio M_T es la torsión que la varilla induce sobre las bolas, pero en realidad las varillas son meros transmisores. Quien realmente genera esta torsión es el cable.

Para entender la situación pensemos en un caso en que la varilla estuviera sujeta a un eje que pudiera girar libremente. Entonces, bastaría dar un pequeño impulso a las bolas (o a una de ellas) para que ambas empezaran a girar, y no se detendrían nunca si no fuera por el rozamiento. Ahora pensemos en una varilla sujeta a una soga gruesa. La situación es completamente distinta. Cuando giramos la varilla (o las bolas) la cuerda se retuerce, y ofrece mayor resistencia cuanto más intentamos retorcerla. Si dejamos de empujarlas, las bolas no siguen girando indefinidamente, sino todo lo contrario, el cable las hace girar en sentido inverso para dejar de estar retorcido. Ahora bien, cuando la varilla vuelva a estar en su posición inicial, con el cable sin retorcer, las bolas estarán moviéndose a cierta velocidad, con lo que no se detendrán en seco y seguirán avanzando, y el cable se retorcerá en sentido contrario mientras trata de frenarlas. Cuando lo consiga, estará retorcido y seguirá empujándolas en sentido contrario, y el resultado es una oscilación que sólo se atenuará por el rozamiento.

Este comportamiento de los cables cuando son retorcidos se cuantifica mediante (la versión angular de) la *ley de Hooke*, que afirma que el momento total que el cable ejerce sobre las bolas a través de la varilla es proporcional al ángulo θ en que ésta se encuentra desplazada respecto de su posición de equilibrio (la posición en la que el cable no está retorcido en absoluto), y su sentido es el inverso al del desplazamiento.

Si tomamos el sistema de referencia de modo que la posición de equilibrio de la varilla es la correspondiente a $\theta_1 = 0$ (y a partir de aquí llamaremos θ a θ_1 y $\theta + \pi$ a θ_2), la ley de Hooke para la balanza de torsión afirma que

$$M_T = -\kappa\theta,$$

para cierta constante κ que depende del material del que está hecho el cable, así como de su geometría (la constante es mayor cuanto más grueso y más corto es el cable). En realidad, como en el caso lineal de los muelles, la ley de Hooke sólo es aplicable cuando el ángulo θ es pequeño. Sólo teniendo esto en cuenta se puede interpretar coherentemente el signo negativo de la igualdad anterior, que expresa que si la varilla está girada en sentido positivo, el momento corresponderá a unas fuerzas que apuntan en sentido negativo, y viceversa.

Así pues, la ecuación que determina el movimiento de una balanza de torsión es

$$LF_t - \kappa\theta = \frac{1}{2}mL^2\theta'', \quad (4.11)$$

siempre y cuando el ángulo θ no sea excesivamente grande.

La constante κ se puede calcular aplicando a la balanza de torsión unas fuerzas conocidas y midiendo el ángulo θ que forma con la posición inicial cuando se estabiliza, pero hay una forma mucho más simple, que consiste en desplazar ligeramente la balanza de su posición de equilibrio y medir el periodo de oscilación. En efecto, la ecuación del movimiento (con $F_t = 0$) será

$$\theta'' = -\frac{2\kappa}{mL^2}\theta,$$

y ya sabemos que la solución de esta ecuación diferencial es

$$\theta = \theta_0 \cos\left(\sqrt{\frac{2\kappa}{mL^2}}t\right),$$

que corresponde a una oscilación de periodo

$$T = 2\pi\sqrt{\frac{mL^2}{2\kappa}},$$

luego

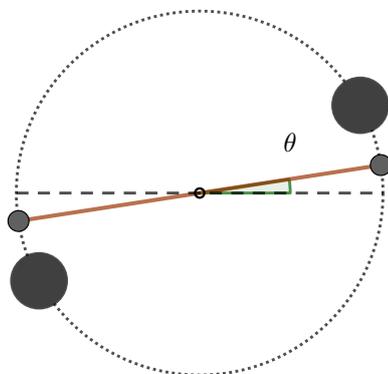
$$\kappa = \frac{2\pi^2mL^2}{T^2}. \quad (4.12)$$

Con esto ya estamos en condiciones de describir el experimento de Cavendish. Consistía en una balanza de torsión formada por dos esferas de plomo, digamos

de 25 mm de radio⁷ sujetas a una barra de madera de 2 m de longitud colgada de un cable en su posición de equilibrio. Luego les puso a una distancia de 25 cm, en la circunferencia sobre la que podía girar la barra y a su misma altura, otras dos esferas de plomo, de 15 cm de radio.

Las esferas grandes estaban fijas, pero las pequeñas podían girar y la pequeña resistencia del cable permitió que la débil atracción gravitatoria entre las bolas cercanas se manifestara. La balanza empezó a oscilar muy lentamente: en las condiciones que hemos descrito y dependiendo de la constante de torsión del cable, la balanza podría tardar unos 7 minutos en detenerse para volver atrás, tras haber recorrido unos 0.007 rad, es decir, 7 mm sobre la circunferencia, lo que supone una velocidad de 1 mm por minuto, aproximadamente.

La figura muestra la balanza desde arriba, pero las medidas no están a escala para que sean apreciables:



Dado que la balanza había sido diseñada para minimizar toda perturbación posible,⁸ incluyendo el rozamiento, las oscilaciones podrían prolongarse un día entero. Para reducir el error, Cavendish midió varias veces la amplitud máxima de cada oscilación para obtener un promedio, y tomó $\theta = 0.007/2 = 0.0035$ como el punto de equilibrio en el cual la balanza debería estabilizarse si se le dejaba tiempo suficiente (pero era mejor no dejar pasar dicho tiempo y quedarse con el valor medio, pues durante ese tiempo podrían acumularse perturbaciones que distorsionaran el resultado).⁹ Más difícil era medir el periodo T de las oscilaciones, pues no es fácil determinar cuándo termina un periodo y empieza el siguiente si la velocidad es de 1 mm por minuto aproximadamente. Aquí el truco está en no tratar de medir el tiempo que tarda en llevarse a cabo un

⁷Los datos que damos son similares a los que consideró Cavendish, pero no tratamos de reproducirlos fielmente porque sus unidades no eran las del sistema métrico decimal y complicarían innecesariamente los cálculos.

⁸Entre otras medidas, Cavendish emplazó su balanza en una habitación sellada contra corrientes de aire, y él mismo se mantuvo fuera de ella durante el experimento. Acercó las bolas fijas con una polea y midió los resultados usando dos pequeños telescopios situados en orificios en las paredes.

⁹Actualmente, en las réplicas del experimento se fija un espejo en a varilla, se le enfoca con un láser y se mide el reflejo en una pantalla situada a varios metros de distancia, de modo que un segundo de arco se corresponde con una longitud sobradamente apreciable.

periodo, sino la suma de los tiempos de varios periodos. Por ejemplo, el tiempo que tardan en producirse 10 oscilaciones es de 2 horas, 27 minutos y 19 segundos, pero no somos capaces de determinar en qué segundo ha terminado la última de ellas y lo dejamos en 2 horas y 27 minutos, el error de 19 segundos en la medición se traduce en un error de 1.9 segundos en la estimación del periodo. Tomemos, pues, $T = (2 \cdot 3600 + 27 \cdot 60)/10 = 882$ s.

Con la determinación de los valores de θ y T , Cavendish había pesado la Tierra. En efecto, si despreciamos las pequeñas variaciones de la fuerza gravitatoria entre las esferas debidas a la oscilación de la balanza, el periodo T es el mismo que se produciría si sobre la balanza no actuara ninguna fuerza tangencial, pero el centro de las oscilaciones no es el punto de equilibrio inicial $\theta_0 = 0$, sino el punto θ donde la balanza puede mantenerse en reposo, es decir, con velocidad angular $\theta'' = 0$, que según la ecuación 4.11, cumple $\kappa\theta = LF_t$. (Esto se razona exactamente igual que vimos en el ejemplo sobre la ley de Hooke (pág. 232) que las oscilaciones de un muelle con un extremo fijo y otro libre tienen el mismo periodo tanto si sobre el objeto sólo actúa la fuerza del muelle como si actúa también su peso constante, si bien el centro de las oscilaciones pasa a ser el punto en el que la fuerza de muelle iguala al peso.) Por lo tanto, dicho periodo T nos permite calcular la constante de torsión del cable mediante (4.12). Por otro lado, la fuerza de atracción entre cada par de esferas es

$$F = \frac{GMm}{r^2},$$

donde la distancia r entre los centros de las bolas se puede calcular con trigonometría a partir de θ . Cavendish despreció la diferencia entre F y la componente tangencial F_t e igualó

$$\kappa\theta = L\frac{GMm}{r^2},$$

con lo que, sustituyendo la expresión (4.12) para κ , obtenemos¹⁰ una fórmula para G :

$$G = \frac{2\pi^2 Lr^2\theta}{MT^2}.$$

Para aplicar esta fórmula a nuestro simulacro de experimento necesitamos

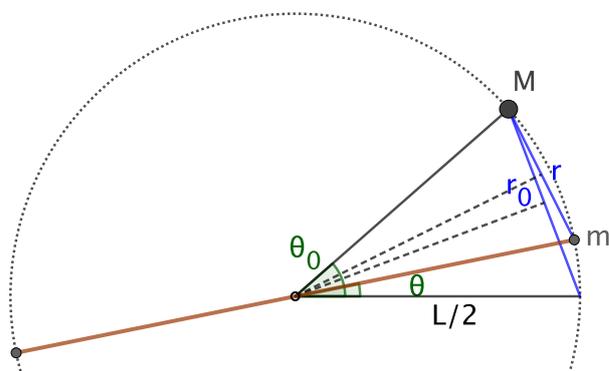
$$r = L \operatorname{sen} \frac{\theta_0 - \theta}{2},$$

donde θ_0 es el ángulo que formaban inicialmente las esferas, que podemos calcular a partir de la distancia inicial $r_0 = 0.25$ m que las separaba:

$$\theta_0 = 2 \operatorname{arcsen} \frac{r_0}{L} = 0.25 \operatorname{rad},$$

de donde $r = 0.246527$ m.

¹⁰Cavendish ni siquiera mencionaba la constante G , sino que la sustituía mediante (4.10) por gR_T^2/M_T , con lo que obtenía la masa de la Tierra M_T .



Por último, teniendo en cuenta que la densidad del plomo es de $\rho = 11\,340 \text{ kg/m}^3$, obtenemos que¹¹

$$M = \frac{4}{3}\pi(0.15)^3\rho = 160.315 \text{ kg}.$$

Finalmente obtenemos

$$G = 6.73 \cdot 10^{-11} \text{ Nm}^2/\text{kg}^2.$$

El valor que obtuvo Cavendish (o el que habría obtenido si hubiera calculado G a partir de sus resultados) fue $G = 6.74 \cdot 10^{-11}$ y el valor aceptado actualmente es $G = 6.67430 \cdot 10^{-11}$, con lo que el error que cometió fue inferior al 1%.

Según (4.10), la masa de la Tierra (con el valor de G correspondiente a los cálculos de Cavendish) resulta ser

$$M_T = \frac{gR_T^2}{G} = \frac{9.8(6\,371\,000)}{6.74 \cdot 10^{-11}} = 5.9 \cdot 10^{24} \text{ kg},$$

que difiere en poco más del 1% del valor aceptado actualmente.

Pero todavía podemos extraer más consecuencias. La tercera ley de Kepler, en su forma (4.9), nos permite calcular ahora la masa del Sol:

$$M_S = \frac{4\pi^2 a^3}{GT^2},$$

donde T es el periodo de revolución de cualquier planeta (por ejemplo la Tierra) y a es el semieje mayor de su órbita. El resultado es

$$M_S = \frac{4\pi^2 149\,500\,000\,000^3}{6.67 \cdot 10^{-11} (365.25 \cdot 24 \cdot 3\,600)^2} = 1.98 \cdot 10^{30} \text{ kg}.$$

■

¹¹En un experimento real sería preferible pesar las bolas en lugar de calcular teóricamente su peso a partir de su densidad.

Caída libre desde gran altura Consideremos el problema siguiente:

¿Cuánto tiempo tardaría la Tierra en caer al Sol si se parara de repente?

Cuando un cuerpo cae en las proximidades de la Tierra (o del Sol), su movimiento está determinado por el hecho de que su aceleración es constante, pero esto sólo es válido en la medida en que podamos desprestigiar la variación paulatina de la intensidad de la gravedad a medida que disminuye la distancia. Cuando no es así, tenemos que recurrir a la fórmula general de la ley de gravitación universal.

En realidad ya hemos considerado este caso, pues en la demostración de la primera ley de Kepler hemos visto que si un cuerpo está en reposo en presencia del campo gravitatorio de un objeto masivo, se moverá hacia él en línea recta de modo que la distancia ρ evoluciona según la ecuación diferencial (4.6):

$$\rho'' = -\frac{GM}{\rho^2},$$

con las condiciones iniciales $\rho(0) = \rho_0$ y $\rho'(0) = 0$.

Aunque en realidad es irrelevante, podemos pensar que $\rho(t)$ es la distancia de la Tierra al Sol en el instante t de su hipotética caída. La velocidad $v = \rho'$ cumple que $v' = -GM/\rho^2 < 0$, lo que significa que la función $\rho(t)$ es estrictamente decreciente en todo momento (como era de esperar). Por consiguiente, podemos considerar la función inversa $t(\rho)$ que nos da el instante t en el que la Tierra estará a distancia ρ del Sol, y a su vez podemos considerar la función compuesta $v(\rho) = v(t(\rho))$ que nos da la velocidad de la Tierra en el (único) momento en que se encuentre a distancia ρ del Sol. La regla de la cadena nos da que

$$\rho'' = \frac{dv}{dt} = \frac{dv}{d\rho} \frac{d\rho}{dt} = \frac{dv}{d\rho} v.$$

Así pues, la ecuación del movimiento puede expresarse como

$$\frac{dv}{d\rho} v = -\frac{GM}{\rho^2}.$$

Así hemos pasado de una ecuación diferencial de segundo orden a una ecuación de primer orden de variables separables:

$$v dv = -\frac{GM}{\rho^2} d\rho.$$

Integrando queda

$$\frac{v^2}{2} = \frac{GM}{\rho} + c.$$

La constante c se calcula a partir de la condición inicial $v(\rho_0) = 0$, de modo que $c = -GM/\rho_0$ y así:

$$\frac{v^2}{2} = \frac{GM}{\rho} - \frac{GM}{\rho_0} = \frac{GM}{\rho_0} \frac{\rho_0 - \rho}{\rho},$$

luego

$$\frac{d\rho}{dt} = v = -\sqrt{\frac{2GM}{\rho_0}} \sqrt{\frac{\rho_0 - \rho}{\rho}}.$$

Nuevamente tenemos una ecuación de variables separables:

$$\sqrt{\frac{\rho}{\rho_0 - \rho}} d\rho = -\sqrt{\frac{2GM}{\rho_0}} dt$$

La integral del primer miembro la hemos calculado en (C.15):

$$\rho_0 \arcsen \sqrt{\frac{\rho}{\rho_0}} - \sqrt{\rho(\rho_0 - \rho)} = -\sqrt{\frac{2GM}{\rho_0}} t + c.$$

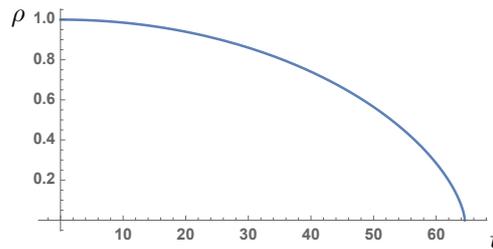
La constante se calcula con la condición inicial $\rho(0) = \rho_0$, que nos da $c = \pi\rho_0/2$. Por lo tanto:

$$\rho_0 \arcsen \sqrt{\frac{\rho}{\rho_0}} - \sqrt{\rho(\rho_0 - \rho)} = -\sqrt{\frac{2GM}{\rho_0}} t + \frac{\pi\rho_0}{2}.$$

Para terminar de resolver la ecuación diferencial tendríamos que despejar ρ en función de t , pero no es posible hacerlo. Es una situación bastante frecuente que al integrar una ecuación diferencial lleguemos a una ecuación que determina implícitamente la solución, pero que no nos permite despejarla explícitamente. Lo que sí que podemos hacer es despejar la función inversa que nos dice el tiempo que tardaría la Tierra en llegar a una posición dada ρ :

$$t = \sqrt{\frac{\rho_0}{2GM}} \left(\frac{\pi\rho_0}{2} + \sqrt{\rho(\rho_0 - \rho)} - \rho_0 \arcsen \sqrt{\frac{\rho}{\rho_0}} \right).$$

Representando esta función con la variable ρ en el eje horizontal tenemos una representación de $\rho(t)$ (donde hemos representado t en días y ρ en unidades astronómicas):



En particular podemos calcular el tiempo que tardará la Tierra en llegar hasta el centro del Sol.

$$t(0) = \sqrt{\frac{\rho_0}{2GM}} \frac{\pi\rho_0}{2} = 64 \text{ días, } 13 \text{ h, } 37 \text{ min, } 82 \text{ s.}$$

En realidad es más realista determinar cuándo impactaría con la superficie del Sol, que es sólo un poco menos. Sólo hay que evaluar t en la longitud del radio del Sol:

$$t(6.957 \cdot 10^8) = 64 \text{ días, } 13 \text{ h, } 25 \text{ min, } 17 \text{ s.} \quad \blacksquare$$

La velocidad de fuga En su novela *De la Tierra a la Luna*, Julio Verne plantea la posibilidad de construir un cañón gigantesco que dispare un proyectil con velocidad suficiente para que no caiga y llegue hasta la luna. *¿Cuál es esa mínima velocidad de fuga que permite escapar a los objetos que se lanzan verticalmente?*

Mientras el objeto ascienda, su velocidad irá disminuyendo, luego podemos considerar como antes la función $v(\rho)$, que nos lleva a la ecuación diferencial

$$\frac{v^2}{2} = \frac{GM_T}{\rho} + c,$$

sólo que ahora la condición inicial es $v(r_T) = v_0$, donde r_T es el radio de la Tierra y v_0 la velocidad de fuga que queremos calcular. Entonces

$$\frac{v^2}{2} = \frac{GM_T}{\rho} + \frac{v_0^2}{2} - \frac{GM_T}{r_T},$$

luego

$$v = \sqrt{\frac{2GM_T}{\rho} + v_0^2 - \frac{2GM_T}{r_T}}.$$

A medida que ρ crece, el término $2GM_T/\rho$ tiende a 0, luego si $v_0^2 < 2GM_T/r_T$, llegará un momento en que la velocidad se anulará, y a partir de ese momento el cuerpo caerá. Por el contrario, si $v_0^2 \geq 2GM_T/r_T$, la velocidad será siempre positiva, luego el cuerpo no caerá. La velocidad de fuga es, pues,

$$v_f = \sqrt{\frac{2GM_T}{r_T}}.$$

Para el caso de la Tierra es

$$v_f = \sqrt{\frac{2 \cdot 6.67430 \cdot 10^{-11} \cdot 5.9736 \cdot 10^{24}}{6.371 \cdot 10^6}} = 11\,187 \text{ m/s}$$

(más de 40 000 km/h.) \blacksquare

Satélites geoestacionarios *¿A qué altura hay que situar un satélite artificial para que siga una órbita circular en el plano ecuatorial terrestre a la misma velocidad angular que gira Tierra?*

Según hemos visto en el apartado sobre movimiento circular uniforme (página 214), la aceleración (normal) de un satélite que gire con velocidad angular

constante ω será $a = \omega^2 r$. En la igualdad $\vec{F} = m\vec{a}$, donde \vec{F} es el peso del satélite y m es su masa. Tanto \vec{F} como \vec{a} apuntan hacia el centro de la Tierra, luego la igualdad vectorial equivale a la igualdad de los módulos:

$$\frac{GM_T m}{(R_t + h)^2} = m\omega^2(R_t + h),$$

luego

$$h = \sqrt[3]{\frac{GM_T}{\omega^2}} - R_t.$$

La velocidad angular de la Tierra es de 2π radianes cada 24 horas, aunque si queremos ser precisos, el tiempo que tarda la Tierra en dar una vuelta completa sobre sí misma no es el tiempo que tarda el Sol en dar una vuelta completa en el cielo (el día solar), sino más bien el tiempo que tarda una estrella en hacerlo (día sideral), que es de 23 h, 56 m, 4.1 s = 86 164.1 s:

$$\omega = \frac{2\pi}{86\,164.1} \text{ rad/s.}$$

Usando $M_T = 5.9736 \cdot 10^{24}$ kg y $R_T = 6\,378.1$ km (consideramos en este caso el radio ecuatorial de la Tierra, que es mayor que el radio medio), el resultado es

$$h = 35\,789 \text{ km.}$$

■

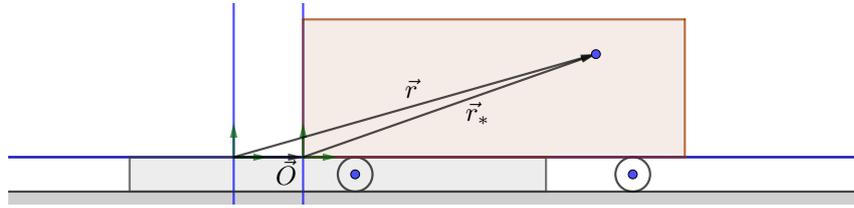
4.4 Sistemas de referencia no inerciales

Vamos a analizar ahora el principio de relatividad de Galileo:

Las leyes del movimiento son exactamente las mismas cuando se expresan respecto de un sistema de referencia en reposo o respecto de un sistema de referencia que se mueve con velocidad constante.

Para analizar este principio tenemos que traer al primer plano un hecho que hasta ahora hemos tenido en cuenta sin darle la importancia debida, y es que toda descripción de un movimiento depende necesariamente de un sistema de referencia prefijado, y en la práctica, un sistema de referencia tiene que estar vinculado necesariamente a un objeto físico de referencia respecto del cual medir la posición de los objetos.

Consideremos dos observadores, uno situado en un vagón de un tren y otro en el andén de la estación. En la práctica, la palabra “observador” se usa como sinónimo de “sistema de referencia”, de modo que cuando hablamos de un observador situado en el andén entendemos que hemos fijado un sistema de referencia respecto del cual el andén está en reposo, digamos con el eje X en la dirección de la vía y el eje Z en vertical, como muestra la figura:



Similarmente, cuando hablamos de un observador situado en el tren entendemos que hemos fijado otro sistema de referencia respecto del cual el tren está en reposo, por ejemplo, con origen de coordenadas en una esquina de un vagón. Cada observador asignará unas coordenadas distintas a un mismo punto. Por ejemplo, si un viajero sostiene una piedra en su mano, el observador en el andén le asignará unas coordenadas $\vec{r} = (x, y, z)$ y el observador en el tren otras $\vec{r}_* = (x^*, y^*, z^*)$, de modo que la relación entre ambas es

$$\vec{r} = \vec{O} + \vec{r}_*,$$

donde \vec{O} es la posición en que el observador en el andén ve el origen del coordenadas del observador del tren. Concretamente, si el tren avanza por la vía a velocidad constante v_O y tomamos como instante $t = 0$ el momento en que los orígenes de ambos sistemas de referencia coinciden, la relación es

$$(x, y, z) = (v_O t, 0, 0) + (x^*, y^*, z^*) = (v_O t + x^*, y^*, z^*).$$

Pero, \vec{r} y \vec{r}_* son funciones del tiempo, y al derivar queda $\vec{v} = \vec{v}_O + \vec{v}_*$ o, explícitamente:

$$(v_x, v_y, v_z) = (v_O + v_x^*, v_y^*, v_z^*),$$

y volviendo a derivar obtenemos la relación entre las aceleraciones: $\vec{a} = \vec{a}_*$. En conclusión la situación es la siguiente:

- Los observadores *difieren superficialmente* sobre cuál es la posición del objeto en cada instante, en el sentido de que la diferencia numérica entre las coordenadas expresa meramente que están observándolo desde posiciones diferentes, pero en realidad están de acuerdo en cuál es la posición del objeto en el espacio. La discrepancia numérica es análoga a la que puede darse entre alguien que afirme que una persona mide 1.80 m y otro que afirme que esa misma persona mide 5.90 pies. Se trata de dos valores numéricos que hacen referencia a una misma longitud.
- Los observadores *difieren sustancialmente* sobre cuál es la velocidad del objeto en cada instante, pues si, por ejemplo, el tren se mueve a 30 m/s y el observador situado en él afirma que el objeto está en reposo, para el observador situado sobre el andén la velocidad será (30, 0, 0) m/s. En este caso, una velocidad de 0 m/s y una velocidad de 30 m/s no son dos expresiones numéricas que hacen referencia a una misma velocidad, sino que son dos percepciones distintas de la velocidad del objeto. En otras palabras, *la velocidad de un móvil es relativa al observador* y la relación entre las velocidades es el *principio de adición de velocidades*: $\vec{v} = \vec{v}_O + \vec{v}_*$.

Si un arco dota a una flecha de una velocidad de 80 m/s, pero el arquero que la dispara se está moviendo (respecto de otro observador) a 330 m/s en la dirección del disparo, la flecha saldrá a una velocidad de 80 m/s para el arquero y a $330 + 80 = 410$ m/s para el segundo observador, luego ambos coincidirán en que la flecha se alejará del arquero a 80 m/s y no le impactará en la cara, en contra de la opinión de los geocentristas coetáneos de Galileo.

- Los observadores *coinciden* sobre cuál es la aceleración del objeto en cada instante. Por ejemplo, si el viajero deja caer la piedra que sostiene, digamos a 1 m de altura en $t = 0$, la única fuerza que actúa sobre ella es su peso, que le provoca una aceleración de $(0, 0, -9.8)$ m/s, tanto para el observador situado sobre el tren como para el del andén.

Así pues, desde el andén, si la posición inicial de la piedra es, digamos, $\vec{r}_0 = (2, 1, 1)$, su movimiento vendrá dado por

$$\vec{r} = \vec{r}_0 + \vec{v}_0 t + \frac{1}{2} \vec{a} t^2 = (2, 1, 1) + (30, 0, 0)t + \frac{1}{2}(0, 0, -9.8)t^2,$$

es decir:

$$\vec{r} = (2 + 30t, 1, 1 - 4.9t^2),$$

que corresponde a una parábola, como sucede siempre que se lanza un objeto con velocidad horizontal no nula. Ahora bien, esta trayectoria, vista desde el tren, es

$$\vec{r}^* = \vec{r} - \vec{O} = (2 + 30t, 1, 1 - 4.9t^2) - (30t, 0, 0) = (2, 1, 1 - 4.9t^2),$$

que corresponde a una recta vertical, que es justo la trayectoria que debe seguir un objeto que se deja caer desde el reposo. En particular, para el observador situado sobre el tren, la piedra empieza a caer desde el punto $\vec{r}_*(0) = (2, 1, 1)$ y llega al suelo en el punto $(2, 1, 0)$, después de haber seguido una trayectoria perfectamente vertical (sin que las dos primeras coordenadas se hayan modificado). No es cierto que, a causa del movimiento del tren, la piedra tenga que caer unos metros hacia atrás del punto desde donde se la ha dejado caer.

Así, un observador situado sobre el tren no puede saber si éste está parado o si se mueve a velocidad constante observando la forma en que cae una piedra, pues en ambos casos la verá caer verticalmente.

Más en general, hemos razonado que siempre que un observador se mueve con velocidad constante respecto de otro, ambos medirán la misma aceleración \vec{a} en cualquier objeto. Si aceptamos que la masa de un objeto tampoco depende del observador (si un objeto tiene 3 kg de masa para un observador, tiene 3 kg de masa para cualquier observador) y que las fuerzas que actúan sobre un objeto también son las mismas para cualquier observador (si un objeto es atraído por la Tierra con una fuerza de 20 N para un observador, lo mismo vale para cualquier

otro observador) concluimos¹² que si un observador ve satisfecha la segunda ley de Newton $\vec{F} = m\vec{a}$, también la verá satisfecha cualquier otro observador que se mueva respecto a él con velocidad constante. De hecho, no es que se seguirá cumpliendo la ecuación globalmente, sino que se cumplirá término a término, es decir, que tanto \vec{F} , como m , como \vec{a} tomarán los mismos valores para ambos.

Por lo tanto, el observador situado en el tren no puede observar nada en el movimiento de los objetos de su entorno que le permita deducir si está parado o en movimiento (con velocidad constante), pues todos ellos se moverán según la segunda ley de Newton en ambos casos. Puede ocurrir —según hemos visto— que uno diga que un objeto se mueve siguiendo una parábola y otro diga que ese mismo objeto se mueve siguiendo una recta, pero ello se debe a que ambos pueden discrepar en la velocidad que tiene el objeto en cada instante, y la fórmula $\vec{F} = m\vec{a}$ proporciona una trayectoria distinta según la velocidad inicial.

Ahora bien, en todo este planteamiento es fundamental que un observador se mueva a velocidad constante respecto del otro. El lector sabrá sin duda lo que sucede si viaja en un tren, o en un coche, cuando éste acelera o frena. Cualquier objeto que no esté sujeto al vehículo se ve impulsado hacia atrás (si el vehículo acelera) o hacia adelante (si frena) sin que la fuerza que siente pueda atribuirse a ninguna causa física. Por poner un ejemplo sencillo: si en el pasillo del vagón del tren hay una pelota en reposo, en el momento en que el tren frene se verá impulsada hacia adelante sin que nadie la haya tocado. Para el observador en el andén, la situación es simple: la pelota no estaba en reposo, sino moviéndose a la misma velocidad que el tren, y cuando éste frena, la fuerza que frena al tren no actúa sobre la pelota, luego ésta continúa moviéndose de acuerdo con el principio de inercia y, si acaba chocando con la pared delantera del vagón, no es porque ella haya sufrido una aceleración, sino porque la pared del vagón “que corría ante ella” ha frenado.

Sin embargo, cuando tratamos de analizar la situación desde el punto de vista de un observador situado sobre el tren nos encontramos con un problema de raíz: la segunda ley de Newton no es aplicable. Las únicas fuerzas que están actuando sobre la pelota son su peso, dirigido hacia abajo, y la reacción eléctrica del suelo, dirigida hacia arriba y que compensa al peso, impidiendo que la pelota rompa el suelo del vagón y caiga a la vía. La suma de las fuerzas que actúan sobre la pelota es nula y sigue siendo así en el momento en que el tren frena, pese a lo cual la pelota acelera (respecto del tren). No hay ningún misterio en ello: sobre el tren está actuando una fuerza que lo frena, pero el tren está en reposo respecto del observador montado en él, lo cual hace lógicamente imposible que perciba como tal la aceleración correspondiente (es lógicamente imposible ver cómo frena un tren que para él está en reposo) y, sin embargo, ese frenazo se materializa en una aceleración de los objetos situados sobre el tren.

¹²Cabe mencionar que esto que estamos aceptando es falso. La teoría de la relatividad implica que dos observadores en movimiento relativo discreparán en la distancia que separa a dos objetos en un instante dado, en el tiempo que transcurre entre dos sucesos, en la masa que tiene un objeto o en las fuerzas que actúan sobre él, pero estas discrepancias sólo son apreciables cuando la velocidad relativa entre los observadores es próxima a la velocidad de la luz, por lo que la mecánica newtoniana es una aproximación válida de la dinámica de los objetos que se mueven a velocidades muy alejadas de la velocidad de la luz.

En la práctica, podemos decir que —desde el punto de vista del observador en el tren— dichos objetos están experimentando una fuerza que los empuja (y esto es más evidente, si cabe, cuando el objeto en cuestión es el propio observador que “se siente” empujado hacia adelante), pero es una fuerza peculiar, porque no puede vincularse a ninguna causa física: la pelota no ha sido atraída por la fuerza de la gravedad, ni es una pelota metálica atraída por un imán, ni nadie le ha pegado una patada: nada ni nadie le ha hecho nada “y sin embargo se mueve”.

Esto significa que un viajero en el tren puede deducir que el tren está en movimiento y, más aún, que está acelerando o frenando, en cuanto observa la aparición de estas fuerzas “misteriosas” a las que no puede atribuir una causa física directa, sino únicamente una explicación indirecta basada en deducir que el tren que está en reposo respecto de su propio sistema de referencia no lo está en realidad.

Quizá el lector pueda considerar que todo esto son sofismas filosóficos como la paradoja de Zenón, y que lo único que sucede es que un observador en movimiento no es de fiar, pues dirá que un tren que se mueve está en reposo y que las casas que ve por la ventanilla están moviéndose cuando están paradas, y si partimos de tales falsedades no es extraño llegar a situaciones absurdas. Sin embargo, la situación no es tan simple. De momento, podemos concluir que hay observadores “fiables” que observarán que los objetos se mueven siguiendo la segunda ley de Newton y observadores “no fiables” para los que la segunda ley de Newton fallará y verán pelotas que empiezan a moverse sin que nadie las toque, ni nada las atraiga o repela por gravedad, electricidad, magnetismo, etc. Concretamente, esto sucede debido a que todo sistema de referencia tiene que estar vinculado a un objeto físico (como un tren, una casa, la Tierra, el Sol, etc.) y cualquier fuerza que actúe sobre dicho objeto de referencia será “invisible” para el observador de forma directa, pues el objeto estará en reposo “por definición”, aunque se plasmará en fuerzas “misteriosas” que actúan sobre los objetos sin que puedan ser vinculadas a causa física alguna. Por lo tanto, cabe esperar que los sistemas de referencia “fiables” sean aquellos vinculados a objetos físicos sobre los que no está actuando ninguna fuerza (o, al menos, sobre los que la fuerza total que actúa sobre ellos es nula), luego, según el principio de inercia, son los sistemas de referencia vinculados a objetos en reposo o que se mueven con velocidad constante.

Ahora bien, ¿cómo podemos saber si un sistema de referencia está en reposo? Por ejemplo, no es cierto que la Tierra esté en reposo, ya que gira sobre sí misma y alrededor del Sol, ni el Sol está en reposo, porque se mueve a través de la Galaxia. . . Newton creyó que podía resolver el problema partiendo del supuesto de que las estrellas están en reposo, de modo que los observadores “fiables” son los que están en reposo absoluto, es decir, los que ven las estrellas en reposo y, de acuerdo con el principio de relatividad de Galileo, los que se mueven con velocidad constante respecto de los sistemas de referencia en reposo absoluto. Sin embargo, los astrónomos no tardaron en comprobar que las estrellas se mueven, y en ausencia de unos objetos claramente en reposo, como Newton creía que eran las estrellas, no tenemos ningún criterio físico que nos permita distinguir

qué está en “reposo absoluto” y qué no. Por ello en 1885 el físico alemán Ludwig Lange introdujo el concepto de “sistema de referencia inercial”, que esencialmente es un sistema de referencia en el cual es válida la segunda ley de Newton, es decir, respecto del cual la aceleración que se observa en un objeto es la dada por la fórmula $\vec{F} = m\vec{a}$, donde \vec{F} es la resultante de las fuerzas que actúan sobre él vinculadas a una causa física identificable (atracción gravitatoria, atracción o repulsión eléctrica, etc.), sin contar “fuerzas misteriosas” que obliguen a concluir que alguna fuerza está actuando sobre el propio sistema de referencia siendo, por tanto, invisible al observador.

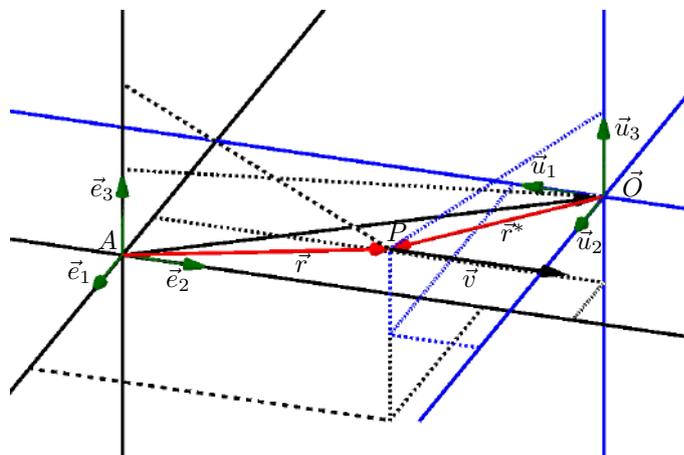
Así, entre los supuestos básicos de la mecánica newtoniana se encuentra el hecho de que existen sistemas de referencia inerciales en los cuales es aplicable la segunda ley de Newton y que, fijado un sistema de referencia inercial, lo son también todos aquellos vinculados a objetos que se mueven respecto de él con velocidad constante. Las leyes de la física son las mismas en cualquier sistema de referencia inercial, por lo que es imposible diseñar cualquier clase de experimento que, a partir de la observación del movimiento de los objetos, nos permita deducir si el objeto al que se vincula un sistema de referencia inercial está en reposo o si se mueve a una determinada velocidad, y mucho menos calcular dicha velocidad. Esto hace que la noción de “reposo absoluto” se vuelva metafísica: si existieran en algún sentido “observadores en reposo absoluto”, serían ciertamente observadores inerciales, pero resultarían físicamente indistinguibles de cualquier otro sistema de referencia inercial, aunque no estuviera en reposo absoluto.

Puede resultar inquietante que, en realidad, el concepto de “sistema de referencia inercial” es una abstracción teórica inalcanzable en la práctica: un sistema de referencia basado en las paredes de una casa no es inercial, porque la casa está girando alrededor de la Tierra y acompañando a ésta en su giro alrededor del Sol, y veremos que esto vuelve inaplicables las leyes de Newton. Ningún sistema de referencia que podamos precisar en la práctica es inercial, por lo que parece que las leyes de Newton son inútiles, ya que nunca se pueden aplicar. En realidad, la situación no es tan desoladora. A continuación vamos a estudiar los sistemas de referencia no inerciales y vamos a ver que las fuerzas “misteriosas”, que técnicamente se llaman “fuerzas ficticias” o “fuerzas de inercia”, no son realmente tan misteriosas, sino que se puede trabajar con ellas exactamente igual que con las fuerzas “reales” y además, en muchos casos resultan prácticamente despreciables, por lo que muchos sistemas de referencia que en teoría no son inerciales (como uno basado en las paredes de una casa) sí que pueden tomarse en la práctica como si lo fueran.

Vamos a empezar comparando dos sistemas de referencia en términos puramente cinemáticos. Como el caso general es bastante sofisticado, vamos a empezar considerando un caso particular, pero un poco más general que el caso del tren que hemos considerado antes. De momento vamos a introducir un elemento que antes no hemos tenido en consideración, y es la posibilidad de que los ejes de coordenadas de ambos observadores no sean paralelos entre sí. Supongamos, pues, que un observador elige un sistema de referencia con origen en un punto A , como muestra la figura, en la que también están representados los

vectores unitarios

$$\vec{e}_1 = (1, 0, 0), \quad \vec{e}_2 = (0, 1, 0), \quad \vec{e}_3 = (0, 0, 1).$$



Por otra parte, supongamos que otro observador elige otro sistema de referencia, cuyo origen tendrá para él coordenadas $(0, 0, 0)$, pero pongamos que para el primer observador está determinado por el vector $\vec{O} = (-1, 5, 1)$. Además, el segundo observador elige los ejes de modo que los vectores $\vec{u}_1, \vec{u}_2, \vec{u}_3$ que para él tienen coordenadas $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ son

$$\vec{u}_1 = (0, -1, 0), \quad \vec{u}_2 = (1, 0, 0), \quad \vec{u}_3 = (0, 0, 1)$$

para el primer observador.

Ahora imaginemos que ambos observadores se fijan en un objeto situado en un punto P , que para el primero de ellos tiene coordenadas $\vec{r} = (3, 4, 2)$, mientras que para el segundo las coordenadas serán $\vec{r}_* = (1, 4, 1)$. La discrepancia en los valores de las coordenadas es sólo aparente. Sería absurdo que los observadores discutieran sobre quién tiene razón, si el que dice que la posición del objeto es $(3, 4, 2)$ o el que dice que la posición es $(1, 4, 1)$. No ha lugar a la discusión porque los dos están diciendo lo mismo en términos distintos.

En efecto, si el primer observador quiere entender lo que está diciendo el segundo cuando dice que la posición del objeto es $\vec{r}_* = (1, 4, 1)$, debe tener en cuenta que esto, para el segundo observador, significa moverse 1 unidad en la dirección de \vec{u}_1 , 4 unidades en la dirección de \vec{u}_2 y 1 unidad en la dirección de \vec{u}_3 , luego, en términos del sistema de referencia del primer observador, lo que dice el segundo es que para llegar al punto P hay que moverse siguiendo el vector

$$\vec{r}_0 = 1\vec{u}_1 + 4\vec{u}_2 + 1\vec{u}_3 = (4, -1, 1).$$

Este vector sigue sin ser $\vec{r} = (3, 4, 2)$, pero ello se debe a que es el vector que indica como llegar a P desde \vec{O} , luego el primer observador debe entender que para llegar a P tiene que moverse hasta \vec{O} y desde allí moverse según \vec{r}_0 . En

resumen, el primer observador debe interpretar la posición $\vec{r}_* = (1, 4, 1)$ como el vector

$$\vec{r} = \vec{O} + \vec{r}_0 = (-1, 5, 1) + (4, 1, 1) = (3, 2, 1),$$

y esto coincide con su propia observación.

Así pues, el hecho de que un observador diga que la posición es $(3, 2, 1)$ y el otro diga que es $(1, 4, 1)$ no significa que discrepen sobre dónde está realmente el cuerpo, sino que significa justamente que ambos lo ven en el mismo lugar, sólo que asignan coordenadas diferentes a ese lugar. En resumen, el lector debe asimilar que tenemos tres vectores distintos:

- $\vec{r} = (x_1, x_2, x_3)$ expresa la posición del objeto respecto del primer observador (el desplazamiento que hay que hacer desde su origen de coordenadas para llegar al objeto).
- $\vec{r}_* = (x_1^*, x_2^*, x_3^*)$ expresa la posición del objeto respecto del segundo observador (el desplazamiento que hay que hacer desde su origen de coordenadas para llegar al objeto).
- \vec{r}_0 expresa el mismo desplazamiento anterior, pero en términos del sistema de referencia del primer observador.

Y la relación entre los tres es que

$$\vec{r} = \vec{O} + \vec{r}_0 = \vec{O} + x_1^* \vec{u}_1 + x_2^* \vec{u}_2 + x_3^* \vec{u}_3.$$

Similarmente, supongamos que el segundo observador determina que la velocidad del objeto en el instante en que pasa por el punto P es $\vec{v}_* = (-2, 0, 0)$. Si el primer observador estuviera de acuerdo en estas cifras, tendríamos un problema, porque ello significaría que no están observando la misma velocidad. En efecto, lo que quiere decir el segundo observador es que el vector velocidad es $\vec{v}_0 = -2\vec{u}_1 + 0\vec{u}_2 + 0\vec{u}_3 = (0, 2, 0)$, luego, si ambos observadores coinciden realmente al medir la velocidad del objeto, el primero tendría que afirmar que la velocidad es $\vec{v} = (0, 2, 0)$ y no $(-2, 0, 0)$. De nuevo la discrepancia es aparente, pues se debe a que —como es evidente a partir de la figura— moverse en sentido negativo respecto al primer eje del segundo sistema de referencia es lo mismo que moverse en sentido positivo respecto al segundo eje del primero. Nuevamente, tenemos tres vectores relacionados con la velocidad:

- $\vec{v} = (v_1, v_2, v_3)$ expresa la velocidad del objeto respecto del primer observador.
- $\vec{v}_* = (v_1^*, v_2^*, v_3^*)$ expresa la velocidad del objeto respecto del segundo observador.
- $\vec{v}_0 = v_1^* \vec{u}_1 + v_2^* \vec{u}_2 + v_3^* \vec{u}_3$ expresa la misma velocidad anterior, pero en términos del sistema de referencia del primer observador.

Nuevamente, la igualdad que tiene que darse si ambos observadores ven moverse al cuerpo a la misma velocidad no es $\vec{v} = \vec{v}_*$, sino $\vec{v} = \vec{v}_0$, ya que el paso de \vec{v}_* a \vec{v}_0 simplemente introduce los cambios debidos al hecho de que ambos observadores contemplan el móvil desde posiciones distintas.

La razón por la que toda esta discusión es relevante y no una mera obviedad complicada con notación matemática es que, como ya hemos visto en el caso del tren, la igualdad $\vec{v} = \vec{v}_0$ no se da en general, sino que puede fallar si un observador está en movimiento respecto del otro.

Por ejemplo, imaginemos que el segundo observador está en un tren que el primero, situado en tierra, ve moverse con velocidad constante $\vec{v}_O = (0, -2, 0)$. Entonces, si este segundo observador ve al objeto moverse a una velocidad constante $\vec{v}_* = (-2, 0, 0)$, eso significa que su velocidad respecto del tren vista por el primer observador es $v_0 = (0, 2, 0)$, pero si el objeto se mueve a velocidad $(0, 2, 0)$ respecto de un tren que se está moviendo a velocidad $(0, -2, 0)$, es decir, si el objeto se mueve sobre el tren a la misma velocidad respecto del tren que éste se está moviendo respecto de la tierra pero en sentido contrario, el observador situado en tierra lo verá parado, por lo que tendremos:

$$\vec{v} = (0, 0, 0) \neq (0, 2, 0) = \vec{v}_0, \quad \vec{v}_* = (-2, 0, 0).$$

Ahora \vec{v}_* es la velocidad del objeto que percibe el observador situado sobre el tren, \vec{v}_0 es esa misma velocidad relativa al tren, pero expresada según el sistema de referencia del observador en tierra, y $\vec{v} = \vec{v}_O + \vec{v}_0 = \vec{0}$ resulta de sumar vectorialmente la velocidad del tren y la velocidad del objeto respecto del tren.

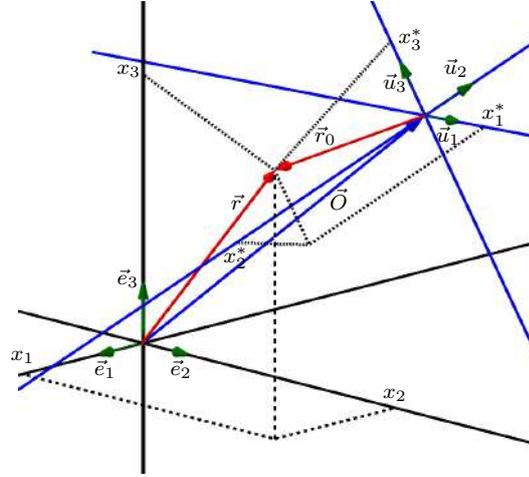
El punto crucial que este ejemplo pretende ilustrar es que el principio de adición de las velocidades no se puede aplicar sumando $\vec{v}_O + \vec{v}_*$, pues cada velocidad está asociada a un sistema de referencia distinto. La segunda coordenada de \vec{v}_O , que es 2, hace referencia a un movimiento en la dirección de \vec{e}_2 , mientras que la segunda coordenada de \vec{v}_* , que es 0 hace referencia a un movimiento en la dirección de $\vec{u}_2 = \vec{e}_1$, luego no tiene sentido sumar esas coordenadas. Para sumar las velocidades, primero tenemos que transformar \vec{v}_* en \vec{v}_0 , de modo que las coordenadas de \vec{v}_O y \vec{v}_0 hagan referencia a desplazamientos en la misma dirección y sentido.

Ahora pasamos ya al caso general, y en él tenemos que contemplar una posibilidad que hasta ahora no hemos considerado, y es que los ejes del segundo sistema de referencia pueden estar en movimiento respecto del primero.

Consideramos, como antes, dos observadores que han fijado sistemas de referencia respectivos, de modo que el origen de coordenadas del segundo está determinado respecto del primero por un vector $\vec{O}(t)$ que depende del tiempo e, igualmente, los vectores que para el segundo observador tienen coordenadas $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ para el primero son $\vec{u}_1(t)$, $\vec{u}_2(t)$, $\vec{u}_3(t)$. En particular, estos vectores son en todo momento unitarios y ortogonales dos a dos. Más aún, suponemos que están bien orientados, en el sentido de que

$$\vec{u}_1 \times \vec{u}_2 = \vec{u}_3, \quad \vec{u}_2 \times \vec{u}_3 = \vec{u}_1, \quad \vec{u}_3 \times \vec{u}_1 = \vec{u}_2. \quad (4.13)$$

(La primera igualdad implica las otras dos.)



Por último, consideramos un móvil puntual cuyo vector de posición es $\vec{r}(t) = (x_1(t), x_2(t), x_3(t))$ para el primer observador y $\vec{r}_*(t) = (x_1^*(t), x_2^*(t), x_3^*(t))$ para el segundo. La figura anterior muestra la situación.

Tal y como hemos discutido previamente, si el primer observador le pregunta al segundo por la posición del móvil y éste responde que está en la posición $\vec{r}_* = (x_1^*, x_2^*, x_3^*)$, el primer observador no debe interpretar esto literalmente, pues las coordenadas están calculadas respecto de unos ejes que no guardan ninguna relación con los suyos. Lo que debe hacer “traducir” esta información pasando al vector

$$\vec{r}_0 = \sum_{i=1}^3 x_i^* \vec{u}_i.$$

y aun así debe tener en cuenta que éste vector sólo informa de como desplazarse desde \vec{O} hasta el móvil, luego el vector de posición respecto del primer sistema de referencia será en realidad:

$$\vec{r} = \vec{O} + \vec{r}_0 = \vec{O} + \sum_{i=1}^3 x_i^* \vec{u}_i. \quad (4.14)$$

Esta ecuación expresa que, para encontrar el móvil siguiendo las indicaciones del segundo observador, hay que desplazarse hasta su origen de coordenadas \vec{O} y, desde allí, desplazarse en la dirección de los vectores \vec{u}_i las distancias (positivas o negativas) x_i^* que el segundo observador dice que son las coordenadas del móvil.

La ecuación anterior relaciona las coordenadas x_i y x_i^* del móvil en ambos sistemas de referencia. Lo que vamos a hacer es derivarla dos veces para obtener la relación entre las velocidades y las aceleraciones, pero antes vamos a definir un vector esencial para expresar el resultado en términos razonablemente simples.

Puesto que cada función $\vec{u}_i \cdot \vec{u}_j$ es constante (igual a 1 o a 0 según si $i = j$ o $i \neq j$), derivando queda que $\vec{u}_i' \vec{u}_j + \vec{u}_i \vec{u}_j' = 0$. Más explícitamente, $\vec{u}_i \vec{u}_i' = 0$, y además podemos definir las funciones

$$\omega_1 = \vec{u}_2' \vec{u}_3 = -\vec{u}_3' \vec{u}_2, \quad \omega_2 = \vec{u}_3' \vec{u}_1 = -\vec{u}_1' \vec{u}_3, \quad \omega_3 = \vec{u}_1' \vec{u}_2 = -\vec{u}_2' \vec{u}_1. \quad (4.15)$$

Definimos también el vector $\vec{\omega} = (\omega_1, \omega_2, \omega_3)$ y

$$\vec{\omega}_0 = \omega_1 \vec{u}_1 + \omega_2 \vec{u}_2 + \omega_3 \vec{u}_3.$$

Notemos que estos vectores dependen únicamente de los sistemas de referencia, pero no del móvil cuyo movimiento queremos estudiar. Enseguida veremos cuál es su interpretación. De momento observemos que $\vec{\omega}_0$ es un vector que puede determinar el primer observador analizando cómo se mueven los ejes del segundo, mientras que $\vec{\omega}$ es el mismo vector “traducido al lenguaje del segundo observador”, pero se da la circunstancia de que éste no puede determinar directamente $\vec{\omega}$, puesto que para él sus ejes de coordenadas están en reposo, luego no aprecia las variaciones que definen a $\vec{\omega}$.

Pasamos ya a derivar la relación (4.14):

$$\vec{v} = \vec{v}_O + \sum_i v_i^* \vec{u}_i + \sum_i x_i^* \vec{u}_i' = \vec{v}_O + \vec{v}_0 + \sum_i x_i^* \vec{u}_i',$$

donde $\vec{v}_O = \vec{O}'$ es la velocidad del origen del sistema de referencia móvil y \vec{v}_0 es el vector que debe interpretar el primer observador si el segundo le dice que la velocidad del móvil es $\vec{v}_* = (v_1^*, v_2^*, v_3^*) = ((x_1^*)', (x_2^*)', (x_3^*)')$. Esta ecuación expresa que la velocidad que percibe el primer observador es la que percibe el segundo (“traducida a su lenguaje” para que la suma tenga sentido), más la velocidad del origen de coordenadas del segundo, más otro término que ahora vamos a analizar y que depende del movimiento de los ejes de coordenadas del segundo observador.

Observemos ahora que todo vector \vec{u} puede expresarse como

$$\vec{u} = a_1 \vec{u}_1 + a_2 \vec{u}_2 + a_3 \vec{u}_3,$$

donde, teniendo en cuenta que los \vec{u}_i son ortogonales dos a dos, al multiplicar por \vec{u}_j resulta que $a_j = \vec{u} \cdot \vec{u}_j$ o, equivalentemente, vemos que todo vector \vec{u} puede expresarse como

$$\vec{u} = \sum_j (\vec{u} \cdot \vec{u}_j) \vec{u}_j.$$

Aplicando esto a los vectores \vec{u}_i' , teniendo en cuenta que $\vec{u}_i' \cdot \vec{u}_i$ y la definición de $\vec{\omega}$, tenemos que

$$\begin{aligned} \sum_i x_i^* \vec{u}_i' &= \sum_{i,j} x_i^* (\vec{u}_i' \cdot \vec{u}_j) \vec{u}_j \\ &= (-x_2^* \omega_3 + x_3^* \omega_2) \vec{u}_1 + (x_1^* \omega_3 - x_3^* \omega_1) \vec{u}_2 + (-x_1^* \omega_2 + x_2^* \omega_1) \vec{u}_3. \end{aligned}$$

Observemos que este vector es la “traducción” al primer sistema de referencia del vector

$$(-x_2^* \omega_3 + x_3^* \omega_2, x_1^* \omega_3 - x_3^* \omega_1, -x_1^* \omega_2 + x_2^* \omega_1) = \vec{\omega} \times \vec{r}_*,$$

es decir, es lo que podemos llamar $(\vec{\omega} \times \vec{r}_*)_0$, pero, más aún, se cumple que $(\vec{\omega} \times \vec{r}_*)_0 = \vec{\omega}_0 \times \vec{r}_0$. En realidad esto es cierto para vectores cualesquiera: Si tenemos dos vectores

$$\vec{c}_* = (c_1^*, c_2^*, c_3^*), \quad \vec{d}_* = (d_1^*, d_2^*, d_3^*),$$

y sus “traducciones”:

$$\vec{c}_0 = c_1^* \vec{u}_1 + c_2^* \vec{u}_2 + c_3^* \vec{u}_3, \quad \vec{c}_0 = d_1^* \vec{u}_1 + d_2^* \vec{u}_2 + d_3^* \vec{u}_3,$$

se cumple la relación

$$(\vec{c}_* \times \vec{d}_*)_0 = \vec{c}_0 \times \vec{d}_0,$$

pues el primer miembro es

$$(c_2^* d_3^* - c_3^* d_2^*) \vec{u}_1 + (c_3^* d_1^* - c_1^* d_3^*) \vec{u}_2 + (c_1^* d_2^* - c_2^* d_1^*) \vec{u}_3$$

y si desarrollamos el segundo miembro:

$$(c_1^* \vec{u}_1 + c_2^* \vec{u}_2 + c_3^* \vec{u}_3) \times (d_1^* \vec{u}_1 + d_2^* \vec{u}_2 + d_3^* \vec{u}_3)$$

usando que $\vec{u}_i \times \vec{u}_i = 0$ y las relaciones (4.13), obtenemos la misma expresión.

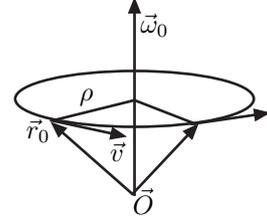
Así pues, hemos llegado a la relación siguiente entre las velocidades:

$$\vec{v} = \vec{v}_O + \vec{v}_0 + \vec{\omega}_0 \times \vec{r}_0. \quad (4.16)$$

Cuando los ejes del segundo observador están en reposo respecto del primero, tenemos $\vec{v} = \vec{v}_O + \vec{v}_0$, que es la expresión general del principio de adición de velocidades, pero en el supuesto de que los ejes estén en movimiento aparece un segundo término que vamos a interpretar. De hecho, de aquí vamos a obtener la interpretación del vector $\vec{\omega}$. Para ello consideramos el caso de una objeto puntual \vec{r} que el segundo observador vea en reposo. Esto significa que $\vec{v}_* = \vec{0}$, luego también $\vec{v}_0 = \vec{0}$ y queda: $\vec{v} - \vec{v}_O = \vec{\omega}_0 \times \vec{r}_0$.

Supongamos también, momentáneamente, que $\vec{v}_O = \vec{0}$, es decir, que el primer observador ve en reposo al origen de coordenadas del segundo (pero ve sus ejes en movimiento). Entonces la relación entre las velocidades se reduce a $\vec{v} = \vec{\omega}_0 \times \vec{r}_0$

Esto significa que los objetos fijos respecto al sistema móvil se mueven con velocidad como indica la figura: tangente a una circunferencia de radio ρ con centro en el eje que pasa por \vec{O} con vector director $\vec{\omega}_0$ y con módulo $v = \omega \rho$, donde $\omega = \|\vec{\omega}_0\|$, de modo que si $\vec{\omega}_0$ permaneciera constante, los puntos en reposo respecto del sistema móvil estarían girando alrededor del eje determinado por \vec{O} y $\vec{\omega}_0$ con velocidad angular ω . Podemos expresar esto diciendo que, en cada instante, el primer observador ve los ejes del segundo observador girando con velocidad angular ω alrededor del eje que pasa por \vec{O} con dirección $\vec{\omega}_0$.



Si no suponemos que \vec{O} es constante, esto no altera la interpretación de $\vec{\omega}_0$: el movimiento del segundo sistema de referencia visto desde el primero puede describirse como una traslación determinada por \vec{O} superpuesta a un giro alrededor del eje determinado por $\vec{\omega}_0$ con velocidad angular ω .

Ahora procedemos a derivar (4.16):

$$\vec{a} = (\vec{v}_O + \vec{v}_0)' + \vec{\alpha}_0 \times \vec{r}_0 + \vec{\omega}_0 \times \vec{r}_0',$$

donde hemos llamado $\vec{\alpha}_0 = \vec{\omega}'_0$. Ahora bien, al pasar de (4.14) a (4.16) hemos visto que la derivada de \vec{r}_0 es $\vec{v}_0 + \vec{\omega}_0 \times \vec{r}_0$, luego

$$\vec{a} = (\vec{v}_0 + \vec{v}_O)' + \vec{\alpha}_0 \times \vec{r}_0 + \vec{\omega}_0 \times \vec{v}_0 + \vec{\omega}_0 \times (\vec{\omega}_0 \times \vec{r}_0).$$

Por último, para derivar

$$\vec{v}_O + \vec{v}_0 = \vec{v}_O + \sum_i v_i^* \vec{u}_i$$

basta observar que se trata de una expresión formalmente idéntica a (4.14), cambiando la función vectorial \vec{O} por \vec{v}_O y las coordenadas x_i^* por v_i^* , luego la derivada es la expresión análoga a (4.16), es decir:

$$(\vec{v}_O + \vec{v}_0)' = \vec{a}_O + \vec{a}_0 + \vec{\omega}_0 \times \vec{v}_0,$$

donde $\vec{a}_O = \vec{v}'_O = \vec{O}''$ es la aceleración con que el primer observador ve moverse al origen de coordenadas O del segundo observador y

$$\vec{a}_0 = \sum_i a_i^* \vec{u}_i$$

es la aceleración que el segundo observador percibe en el móvil expresada en el sistema de referencia del primer observador. En definitiva:

$$\vec{a} = \vec{a}_O + \vec{a}_0 + \vec{\omega}_0 \times (\vec{\omega}_0 \times \vec{r}_0) + 2\vec{\omega}_0 \times \vec{v}_0 + \vec{\alpha}_0 \times \vec{r}_0.$$

Vamos a resumir en un teorema lo que hemos obtenido:

Teorema 4.5 *Consideremos dos sistemas de referencia, de modo que el origen de coordenadas del segundo tenga coordenadas \vec{O} respecto del primero y los ejes de coordenadas del segundo vengan determinados por los vectores $\vec{u}_1, \vec{u}_2, \vec{u}_3$. Sea $\vec{\omega} = (\omega_1, \omega_2, \omega_3)$, donde*

$$\omega_1 = \vec{u}'_2 \vec{u}_3 = -\vec{u}'_3 \vec{u}_2, \quad \omega_2 = \vec{u}'_3 \vec{u}_1 = -\vec{u}'_1 \vec{u}_3, \quad \omega_3 = \vec{u}'_1 \vec{u}_2 = -\vec{u}'_2 \vec{u}_1.$$

Sea

$$\vec{\omega}_0 = \sum_i \omega_i \vec{u}_i.$$

Supongamos que un mismo móvil es observado respecto de ambos sistemas de referencia con posición (x_1, x_2, x_3) y (x_1^, x_2^*, x_3^*) , velocidad (v_1, v_2, v_3) y (v_1^*, v_2^*, v_3^*) y aceleración (a_1, a_2, a_3) y (a_1^*, a_2^*, a_3^*) , respectivamente. Sean*

$$\vec{r}_0 = \sum_i x_i^* \vec{u}_i, \quad \vec{v}_0 = \sum_i v_i^* \vec{u}_i, \quad \vec{a}_0 = \sum_i a_i^* \vec{u}_i$$

la posición, la velocidad y la aceleración respecto del segundo sistema de referencia expresadas en términos del primero. Entonces

$$\vec{r}_0 = \vec{r} - \vec{O}, \quad \vec{v}_0 = \vec{v} - \vec{v}_O - \vec{\omega}_0 \times \vec{r}_0,$$

$$\vec{a}_0 = \vec{a} - \vec{a}_O - \vec{\omega}_0 \times (\vec{\omega}_0 \times \vec{r}_0) - 2\vec{\omega}_0 \times \vec{v}_0 - \vec{\alpha}_0 \times \vec{r}_0,$$

donde \vec{v}_O y \vec{a}_O son la velocidad y la aceleración del origen de coordenadas del segundo sistema de referencia observadas desde el primero y $\vec{\alpha}_0 = \vec{\omega}'_0$.

Para analizar estas fórmulas es preferible pasar ya a términos dinámicos en lugar de puramente cinemáticos. Para ello tenemos que aceptar, como uno de los principios fundamentales de la mecánica newtoniana, que dos observadores cualesquiera coinciden cuando determinan la fuerza “real” que está actuando sobre un objeto en un momento dado, es decir, la suma de todas las fuerzas vinculables a causas físicas (gravedad, electricidad, etc.). Esto significa que si dicha fuerza es $\vec{F} = (F_1, F_2, F_3)$ para un observador y $\vec{F}_* = (F_1^*, F_2^*, F_3^*)$ para otro, no se dará la relación ingenua $\vec{F} = \vec{F}_*$, sino que hay que tener en cuenta las diferencias puramente geométricas entre ambos sistemas de referencia, por lo que la relación correcta es

$$\vec{F} = \vec{F}_0 = F_1^* \vec{u}_1 + F_2^* \vec{u}_2 + F_3^* \vec{u}_3.$$

Esto vale para cada fuerza en particular que actúe sobre el objeto, y también para la fuerza resultante de sumarlas todas. Si ahora suponemos que el primer observador es inercial, con lo que se cumple la relación $\vec{F} = m\vec{a}$, donde m es la masa del objeto, resulta que $\vec{F}_0 = m\vec{a}$, luego multiplicando por m la tercera fórmula del teorema anterior obtenemos

$$m\vec{a}_0 = \vec{F}_0 - m\vec{a}_O - m\vec{\omega}_0 \times (\vec{\omega}_0 \times \vec{r}_0) - 2m\vec{\omega}_0 \times \vec{v}_0 - m\vec{\alpha}_0 \times \vec{r}_0.$$

Podemos expresar esta ecuación exclusivamente en términos del segundo sistema de referencia si expresamos

$$\vec{a}_O = a_{O1}^* \vec{u}_1 + a_{O2}^* \vec{u}_2 + a_{O3}^* \vec{u}_3 = (\vec{a}_O^*)_0, \quad \vec{\alpha}_0 = \alpha_1^* \vec{u}_1 + \alpha_2^* \vec{u}_2 + \alpha_3^* \vec{u}_3,$$

de modo que los vectores \vec{a}_O^* y $\vec{\alpha}_*$ son vectores expresados en el segundo sistema de referencia, pero que sólo pueden medirse directamente desde el primero, ya que dependen del movimiento del segundo sistema de referencia respecto del primero. Así tenemos que

$$m\vec{a}_0 = \vec{F}_0 - m(\vec{a}_O^*)_0 - m\vec{\omega}_0 \times (\vec{\omega}_0 \times \vec{r}_0) - 2m\vec{\omega}_0 \times \vec{v}_0 - m\vec{\alpha}_0 \times \vec{r}_0,$$

e igualando los coeficientes de cada vector \vec{u}_i en ambos miembros (técnicamente, multiplicando por cada \vec{u}_i), obtenemos la igualdad:

$$m\vec{a}_* = \vec{F}_* - m\vec{a}_O^* - m\vec{\omega} \times (\vec{\omega} \times \vec{r}_*) - 2m\vec{\omega} \times \vec{v}_* - m\vec{\alpha}_* \times \vec{r}_*, \quad (4.17)$$

que es el sucedáneo de la segunda ley de Newton que se cumple respecto del segundo sistema de referencia.

Antes hemos dicho que si un sistema de referencia es inercial, también lo serán todos aquellos vinculados a objetos que se muevan respecto a él con velocidad constante. Sin embargo, ahora vemos que es necesario hacer una precisión que antes no habíamos tenido en cuenta porque no nos habíamos planteado la posibilidad de que los ejes del segundo sistema de referencia estuvieran en movimiento:

Dado un sistema de referencia inercial, cualquier otro sistema de referencia cuyo origen de coordenadas se mueva a velocidad constante respecto del sistema dado y cuyos ejes permanezcan en todo momento paralelos a sí mismos (es decir, que no giren) es también un sistema de referencia inercial.

En efecto, la primera condición equivale a $\vec{a}_O^* = \vec{0}$ y la segunda a que los vectores \vec{u}_i sean constantes, lo que a su vez implica que $\vec{\omega} = \vec{0}$, y esto a su vez implica que $\vec{\alpha} = \vec{0}$, por lo que la ecuación anterior se reduce a $\vec{F}_* = m\vec{a}_*$, y esto significa que el segundo sistema de referencia satisface también la segunda ley de Newton.

En cambio, si el primer observador ve acelerar al segundo o ve girar a sus ejes, éste ya no es inercial, sino que percibirá fuerzas ficticias sobre los objetos que no estarán vinculadas a causas físicas que actúen sobre ellos, sino que serán la manifestación indirecta del movimiento del observador, que es directamente “invisible” para él. Lo relevante es que el teorema anterior nos da una descripción matemática completa de dichas fuerzas ficticias, lo que nos permite trabajar con ellas exactamente igual que con las fuerzas “reales”:

Teorema 4.6 *Consideremos dos sistemas de referencia, de modo que el primero sea inercial y el origen de coordenadas del segundo tenga coordenadas \vec{O} respecto del primero y sus ejes de coordenadas vengán determinados por los vectores $\vec{u}_1, \vec{u}_2, \vec{u}_3$. Sea $\vec{\omega} = (\omega_1, \omega_2, \omega_3)$, donde*

$$\omega_1 = \vec{u}'_2 \vec{u}_3 = -\vec{u}'_3 \vec{u}_2, \quad \omega_2 = \vec{u}'_3 \vec{u}_1 = -\vec{u}'_1 \vec{u}_3, \quad \omega_3 = \vec{u}'_1 \vec{u}_2 = -\vec{u}'_2 \vec{u}_1.$$

Sean

$$\vec{\omega}_0 = \sum_i \omega_i \vec{u}_i, \quad \vec{\alpha}_0 = (\vec{\omega}_0)' = \sum_i \alpha_i \vec{u}_i, \quad \vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3).$$

Finalmente, sea

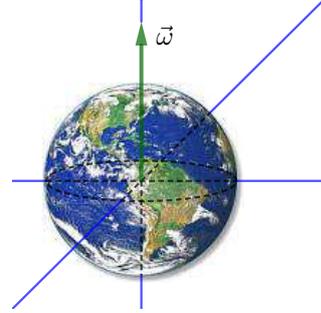
$$\vec{O}'' = \sum_i a_{O_i} \vec{u}_i, \quad \vec{a}_O = (a_{O1}, a_{O2}, a_{O3}).$$

Entonces el segundo sistema de referencia cumple la segunda ley de Newton si a la hora de calcular la fuerza total que actúa sobre un objeto de masa m no sólo sumamos las fuerzas “reales” vinculadas a causas físicas que le afectan, sino también las “fuerzas ficticias” siguientes:

1. La fuerza inercial lineal: $\vec{f}_O = -m\vec{a}_O$.
2. La fuerza centrífuga: $\vec{f}_{\text{cent}} = -m\vec{\omega} \times (\vec{\omega} \times \vec{r})$.
3. La fuerza de Coriolis: $\vec{f}_{\text{Cor}} = -2m\vec{\omega} \times \vec{v}$.
4. La fuerza de Euler: $\vec{f}_{\text{Eu}} = -m\vec{\alpha} \times \vec{r}$.

En efecto, las cuatro fuerzas que hemos enumerado son los términos adicionales del miembro derecho de la ecuación (4.17), salvo que hemos modificado la notación suprimiendo el asterisco que allí distingue a las magnitudes relativas al segundo sistema de referencia (que ahora es innecesario ya que todas las magnitudes son relativas a él).

Vamos a analizar estas cuatro fuerzas en general y en varios ejemplos concretos, entre los cuales consideraremos para todas ellas el caso de un sistema de referencia vinculado a la Tierra, por ejemplo que tenga por ejes las intersecciones entre las paredes y el suelo de una esquina de una habitación de una casa. Ahora bien, en la práctica podemos pasar a un sistema de referencia con origen en el centro de la Tierra, con el eje Z en la dirección de su eje de rotación y con los ejes X, Y pasando por puntos fijos en el ecuador de longitud fija. Si lo comparamos con el asociado a la casa, sucede que uno está en reposo respecto del otro, por lo que las únicas diferencias entre las medidas de posición, velocidad, etc. respecto de uno y otro serán las puramente geométricas, pero en ambos sistemas de referencia tendremos las mismas fuerzas de inercia, aunque se expresen con vectores distintos.

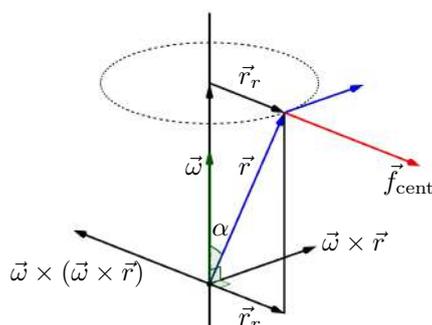


Respecto de un observador situado en la Tierra en reposo respecto a ella, los ejes de coordenadas están fijos, pero respecto de un observador inercial que vea la Tierra “desde lejos”, verá que los ejes X, Y están girando alrededor del eje Z a razón de 2π radianes cada 24 horas, es decir, con velocidad angular $\omega = 2\pi/(24 \cdot 3600)$ rad/s. El vector $\vec{\omega}$ considerado en el teorema anterior apunta en la dirección del eje de giro, luego es en este caso $\vec{\omega} = (0, 0, \omega)$. Notemos que $\omega > 0$ supone que el semieje Z positivo pasa por el polo Norte, pues si miramos la Tierra desde el polo Norte su sentido de giro es positivo: de Oeste a Este (por eso el Sol parece moverse en el cielo de Este a Oeste). Notemos que en la figura $\vec{\omega}$ no está representado a escala, pues su módulo es de unas 73 micras.

La tabla siguiente contiene toda la información necesaria para los cálculos relativos a este ejemplo: las masas de la Tierra, el Sol y la Luna, el radio de la Tierra, las distancias a la Tierra del Sol y la Luna, la constante de gravitación y la velocidad angular de rotación de la Tierra:

M_T	$5.9736 \cdot 10^{24}$ kg	r_T	$6.371 \cdot 10^6$ m
M_S	$1.9891 \cdot 10^{30}$ kg	R_S	$1.496 \cdot 10^{11}$ m
M_L	$7.349 \cdot 10^{22}$ kg	R_L	$3.844 \cdot 10^8$ m
G	$6.67430 \cdot 10^{-11}$ Nm ² /kg ²	ω	$7.2722 \cdot 10^{-5}$ s ⁻¹

La fuerza centrífuga Respecto de un sistema de referencia está girando alrededor del eje señalado por el vector $\vec{\omega}$, cada objeto está sometido a una fuerza centrífuga $\vec{f}_{\text{cent}} = -m\vec{\omega} \times (\vec{\omega} \times \vec{r})$.



Podemos descomponer $\vec{r} = \vec{r}_a + \vec{r}_r$, donde \vec{r}_a es su componente axial (en la dirección de $\vec{\omega}$) y \vec{r}_r es su componente radial, como muestra la figura. Notemos que $\rho = \|\vec{r}_r\| = r \sin \alpha$, donde α es el ángulo que forman \vec{r} y $\vec{\omega}$. Por lo tanto, $\vec{\omega} \times \vec{r} = \vec{\omega} \times \vec{r}_a$ es ortogonal a $\vec{\omega}$ y a \vec{r}_a y su módulo es $\omega \rho$. A su vez, $\vec{\omega} \times (\vec{\omega} \times \vec{r})$ es ortogonal a los dos factores, luego tiene la dirección de \vec{r}_a , pero la regla de la mano derecha muestra que su sentido es el opuesto. Además su módulo es $\omega^2 \rho$, luego es la aceleración normal que tiene un objeto que gira alrededor de un eje a una distancia ρ con velocidad angular ω , luego $f_{\text{cent}} = m\omega^2 \rho$ es el módulo de la fuerza necesaria para que un objeto de masa m gire con radio ρ y velocidad angular ω . Dicha fuerza tiene que apuntar hacia el centro de giro, luego es precisamente $-\vec{f}_{\text{cent}}$. Así pues:

La fuerza centrífuga que experimenta un objeto es la opuesta a la fuerza a la que debe someterse para que gire respecto del eje determinado por $\vec{\omega}$ con velocidad angular ω .

Por ejemplo, imaginemos que un hombre de masa m está de pie sobre un tiiovivo en movimiento que gira con velocidad angular ω y está sujeto a una barra situada a una distancia ρ del eje de giro. Desde el punto de vista de un observador inercial, situado fuera del tiiovivo, sobre el hombre están actuando tres fuerzas: su peso, que se cancela con la reacción del suelo del tiiovivo, y una fuerza que lo sujeta al tiiovivo, ejercida a través del contacto con la barra y del rozamiento con el suelo de sus zapatos. Esta fuerza apunta hacia el eje de giro del tiiovivo y tiene módulo $m\omega^2 \rho$, de modo que se cumple la segunda ley de Newton, pues la aceleración a la que está sometido es una aceleración normal que apunta hacia el eje de giro y tiene módulo $\omega^2 \rho$.

Aquí es fundamental comprender que en un sistema de referencia inercial no hay fuerzas ficticias y, en particular, *no hay ninguna fuerza centrífuga*. Dejando de lado el peso que se cancela, la única fuerza que está actuando sobre el hombre apunta hacia el centro de giro, y es la que lo hace girar. Sin ella, si el hombre se soltara de la barra, abandonaría su trayectoria circular para moverse en línea recta, con lo que se vería expulsado del tiiovivo.

Por el contrario, si describimos la situación respecto de un sistema de referencia no inercial para el que el hombre y el tiiovivo estén en reposo, el hombre no “se siente” acelerado, pues “se siente” en reposo, pero siente igualmente que necesita ejercer una fuerza de intensidad $m\omega^2 \rho$ sobre la barra para sujetarse a

ella o, mejor dicho, para que la barra lo sujete con la reacción correspondiente y compense así una fuerza centrífuga que “siente” que lo empuja hacia afuera con la misma intensidad.

En resumen: desde un sistema de referencia inercial, sobre el hombre actúa una fuerza neta no nula de intensidad $m\omega^2\rho$ (ejercida por la barra a la que se sujeta) que produce en él la aceleración normal necesaria para girar con el tiiovivo, mientras que desde el propio tiiovivo el hombre está en reposo, luego sobre él se cancelan dos fuerzas: una fuerza centrífuga de intensidad $m\omega^2\rho$ que apunta en la dirección opuesta al eje de giro y otra de la misma intensidad ejercida por la barra que la compensa.

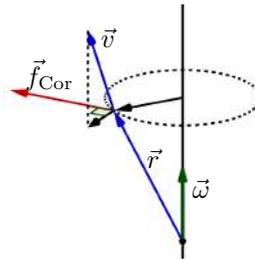
Vamos a calcular ahora la intensidad de la fuerza centrífuga que experimenta un objeto situado en la superficie de la Tierra. Su distancia al eje de giro es $\rho = R_T \cos \alpha$, donde α es la latitud, luego la intensidad de la fuerza centrífuga es

$$\vec{f}_{\text{cent}} = m\omega^2 r_T \cos \alpha = 0.3369m \cos \alpha,$$

que es máxima en el ecuador y nula en los polos, donde equivale al 0.34% del peso del objeto. En otras palabras, la fuerza centrífuga hace que, en el ecuador, un objeto de 1 kg de peso se comporte como si pesara 0.9966 kg.

La fuerza de Coriolis La fuerza de Coriolis $\vec{f}_{\text{Cor}} = -2m\vec{\omega} \times \vec{v}$ recibe este nombre por un trabajo del ingeniero francés Garpard Coriolis publicado en 1835, si bien había sido calculada por Euler en 1749 y su efecto ya había sido considerado en 1661 por los astrónomos Giovanni Battista Riccioli y Francesco Maria Grimaldi en conexión con la balística.

Observamos que \vec{f}_{Cor} es ortogonal a $\vec{\omega}$, luego está contenido en el plano perpendicular al eje de giro que pasa por la posición del objeto, y también es perpendicular a la velocidad \vec{v} , o a su proyección en dicho plano, que es de la que depende realmente. La fuerza de Coriolis es nula cuando la velocidad tiene la dirección del eje de giro y es máxima cuando es ortogonal a ella, y entonces su módulo es $2m\omega v$.

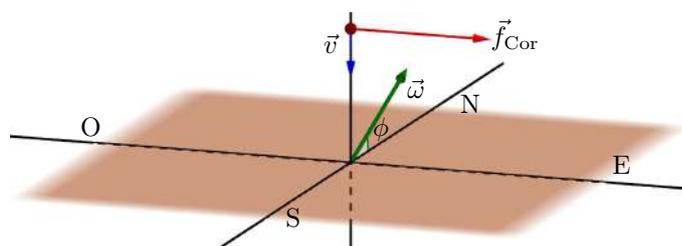


Para interpretar esta fuerza consideremos de nuevo el hombre montado en el tiiovivo, pero supongamos que no permanece en reposo (respecto al tiiovivo), sino que da un paso radialmente hacia afuera. Desde el punto de vista de un observador inercial, si el hombre mueve su pie derecho, resulta que éste parte de un punto en el que se estaba moviendo con el tiiovivo a una velocidad angular ω ,

luego a una velocidad lineal $v_0 = \omega\rho_0$, donde ρ_0 es la distancia al eje. Pero a medida que su pie avanza radialmente hacia afuera, digamos hasta tocar el suelo a una distancia ρ_1 del eje, pasa sobre puntos del suelo que están girando a la misma velocidad angular ω , luego a velocidad lineal $\omega\rho$, que es cada vez mayor, a medida que el pie se aleja del eje, pero la velocidad lineal de su pie nunca deja de ser v_0 (hasta que no toca el suelo de nuevo), por lo que, aunque el hombre pretendiera mover su pie radialmente —y suponiendo que eso es lo que ha hecho desde el punto de vista del observador inercial—, resulta que el tiiovivo corre cada vez más rápido bajo su pie, y termina posándolo un poco hacia atrás respecto del sentido de giro. Desde el punto de vista del hombre, para el que el tiiovivo está en reposo, habrá notado que “algo” ha empujado su pie hacia su derecha (si el tiiovivo gira en sentido positivo visto desde arriba). Ese “algo” es la fuerza de Coriolis.

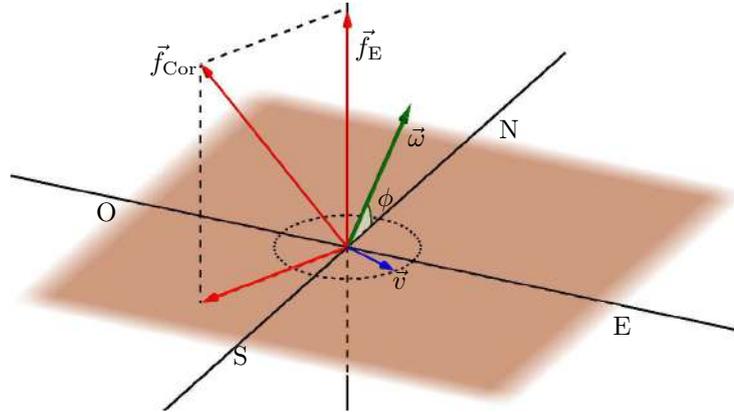
Si aplicamos esto al movimiento de rotación de la Tierra, resulta que los adversarios geocentristas de Galileo tenían razón en que si dejamos caer una piedra desde una torre, la rotación terrestre tendría que notarse en que la trayectoria no es vertical, sino que la piedra tiene que sufrir un desvío, pero no hacia el oeste, como pretendían, sino hacia el este, pues, desde el punto de vista de un observador inercial, a medida que descende pasa por puntos que giran con menor velocidad lineal, luego llega al suelo girando a más velocidad que los puntos a esa altura, luego los habrá adelantado en su giro hacia el este, y habrá caído más hacia el este que el punto desde el que fue lanzada.

Alternativamente, en un punto de la superficie terrestre, el vector $\vec{\omega}$ apunta hacia el Norte, con una inclinación ϕ igual a la latitud del lugar. En el hemisferio norte apunta hacia el cielo, mientras que en el hemisferio sur se mete bajo tierra. Tanto en un caso como el otro, si un cuerpo cae, con lo que su velocidad apunta hacia abajo, la regla de la mano derecha muestra que $\vec{\omega} \times \vec{v}$ apunta siempre hacia el oeste, luego \vec{f}_{Cor} apunta hacia el este. Ahora bien, la figura no está en absoluto a escala, pues, según hemos visto, $\vec{\omega}$ apenas mide unas micras, luego la mayor intensidad que puede tener la fuerza de Coriolis (en el ecuador) es de $2m\omega v = 0.0001454mv$, lo que supone un 0.0015% del peso del objeto que cae por cada m/s de velocidad.



Por ejemplo, si dejamos caer en Florencia (a unos 43° de latitud Norte) una piedra de 1 kg desde 30 m de altura, será empujada hacia el este en su caída por una fuerza de Coriolis de intensidad $2\omega \cos(43^\circ)v = 0.000106371v$ N. Su peso $-g$ hará que llegue al suelo en 2.47 s, y su velocidad vertical en cada instante

será $v_z = -gt$, despreciando la diferencia entre v y v_z , la fuerza horizontal que actúa sobre la piedra en cada instante es $f_{\text{Cor}} = 2\omega \cos(43^\circ)gt$, luego su velocidad horizontal en cada instante será $v = \omega \cos(43^\circ)gt^2$ y el espacio horizontal recorrido será $e = \omega \cos(43^\circ)gt^3/3$, que para $t = 2.47$ nos da un desplazamiento de $e = 2.6$ mm.



Otro caso destacado es el de un objeto que se mueve sobre la superficie terrestre, es decir, tangente a ella. Respecto de un sistema de referencia con el eje X orientado al este y el eje Y orientado al norte, podemos expresar

$$\vec{\omega} = (0, \omega \cos \phi, \omega \sin \phi),$$

donde ϕ es la latitud, y si la velocidad es $\vec{v} = (v_e, v_n, 0)$, un simple cálculo nos da que

$$\vec{f}_{\text{Cor}} = -2m\vec{\omega} \times \vec{v} = 2m\omega \sin \phi (v_n, -v_e, 0) + 2mv_e\omega \cos \phi (0, 0, 1).$$

Así hemos descompuesto la fuerza de Coriolis en una componente tangencial y una componente vertical. La componente vertical tiene intensidad

$$f_E = 2mv_e\omega \cos \phi,$$

entendiendo que apunta hacia arriba si $v_e > 0$ y hacia abajo si $v_e < 0$.

Esta componente vertical de la fuerza de Coriolis recibe el nombre de *fuerza de Eötvös*, por el barón húngaro Loránd Eötvös de Vásárosnamény, quien predijo que la gravedad terrestre medida sobre un barco que navegara hacia el este sería ligeramente menor que medida en un barco que navegara hacia el oeste y, en efecto, poco después, en 1908, se comprobó experimentalmente con dos barcos en el mar Negro que navegaron en sentidos opuestos.

Por ejemplo, si los dos barcos viajaban a 10 m/s en sentidos opuestos, el efecto de la fuerza de Eötvös sobre la gravedad g sería de una variación de ± 0.001 N/kg, lo que se traduce en que si se hace oscilar un péndulo de 5 m de longitud en cada barco y se cronometran 25 oscilaciones, la diferencia de tiempos será de 0.012 s.

En general el efecto de la fuerza de Eötvös es totalmente despreciable porque no puede compararse con el peso del objeto, pero la componente tangencial no se encuentra compensada con ninguna otra fuerza (salvo el rozamiento del aire) y puede hacerse patente. Observemos que el vector $(v_n, -v_e)$ apunta siempre hacia la derecha de \vec{v} , pero $\sin \phi$ es positivo en el hemisferio norte y negativo en el hemisferio sur, lo que se traduce en que la componente tangencial de la fuerza de Coriolis apunta hacia la derecha del movimiento en el hemisferio norte y hacia la izquierda en el hemisferio sur, y su intensidad es

$$2m\omega v \sin \phi.$$

Por lo tanto, la fuerza de Coriolis empuja a los objetos para que giren en sentido horario en el hemisferio norte y en sentido antihorario en el hemisferio sur. Esto se refleja en el movimiento a gran escala de las corrientes de agua y de aire (tornados, ciclones, etc.), que giran en sentidos opuestos según el hemisferio en que se forman.

La fuerza de Euler El ejemplo más cotidiano de fuerza de Euler es la fuerza que experimenta un pasajero de un tiovivo cuando éste arranca o frena. Cuando arranca, se ve empujado hacia atrás, y cuando frena hacia adelante. La explicación es la misma que se aplica a un coche que acelera o frena, pero en el caso del tiovivo no se trata de la fuerza inercial lineal, ya que, considerando un sistema de referencia que gira con el tiovivo con origen de coordenadas en su eje, este origen permanece en reposo también para un observador inercial. Lo que sucede es que $\vec{\omega}$ cambia su módulo (pero no su dirección, pues el eje de giro es siempre el mismo), de modo que $\vec{\alpha} = \dot{\vec{\omega}}$ tiene la misma dirección que $\vec{\omega}$, con el mismo sentido si el tiovivo acelera y sentido opuesto si frena. Esto se traduce en que $\vec{f}_{Eu} = -m\vec{\alpha} \times \vec{r}$, siendo ortogonal a $\vec{\omega}$ y a \vec{r} , apunta en sentido opuesto al giro cuando el tiovivo acelera y en el sentido del giro cuando frena, como tenía que ser.

Si suponemos que el eje de rotación de la Tierra permanece constante (respecto de un observador inercial), entonces la rotación terrestre no genera ninguna fuerza de Euler, pero en realidad, ya en el siglo II a.C., Hiparco, comparando con mediciones antiguas, constató que el polo norte celeste describe se mueve circularmente, avanzando al menos 1° al siglo, lo que supone una vuelta completa en a lo sumo 36 000 años, este movimiento se traduce en que los equinoccios se producen cada año un poco antes que el año anterior, por lo que se denomina movimiento de precesión de los equinoccios. Actualmente se estima que el periodo de precesión es de unos 26 000 años, luego la velocidad angular de la precesión es

$$\Omega = \frac{2\pi}{26\,000 \cdot 365.25 \cdot 24 \cdot 3600} = 7.657 \cdot 10^{-11} \text{ rad/s}.$$

Desde un sistema inercial adecuado, el movimiento de $\vec{\omega}$ viene dado por

$$\vec{\omega} = (\omega \cos \theta \cos \Omega t, \omega \cos \theta \sin \Omega t, \omega \sin \theta),$$

donde $\theta = 23.4^\circ$ es la inclinación de la eclíptica, es decir, el ángulo que forma el eje de rotación de la Tierra con la perpendicular al plano de su órbita alrededor del Sol. Derivando se obtiene que la aceleración angular $\vec{\alpha}$ tiene módulo $\omega\Omega \sin \theta$, por lo que la intensidad máxima de la fuerza de Euler resulta ser de

$$\omega\Omega r_T \sin \theta = 1.41 \cdot 10^{-9} \text{ N/kg},$$

que es menos del 0.000000014% del peso de un objeto, lo que la hace completamente indetectable.

Mareas Consideramos finalmente la fuerza inercial lineal, es decir, la típica fuerza que empuja constantemente a los pasajeros de los medios de transporte. Ya la hemos discutido sobradamente, pero en el caso de la Tierra tiene una manifestación un tanto insólita.

Si observamos nuestro sistema de referencia con origen O en el centro de la Tierra desde un sistema de referencia inercial, su aceleración será $\vec{a}_O = \vec{F}/M_T$, donde M_T es la masa de la Tierra y \vec{F} es la suma de todas las fuerzas que actúan ella. Por consiguiente, la fuerza de inercia lineal que actúa sobre un objeto de masa m es

$$\vec{f} = \frac{m}{M_T} \vec{F}.$$

De entre todas las fuerzas que actúan sobre la Tierra, las más relevantes son la atracción gravitatoria solar y la lunar. Aunque el Sol tiene mucha más masa que la luna, la gravitación lunar resulta ser más relevante por su mayor proximidad, así que vamos a estudiar con detalle la fuerza de inercia que determina sobre un objeto situado en la superficie terrestre.

Ahora nos encontramos con una situación peculiar, porque las fuerzas que frenan un coche no actúan sobre sus pasajeros, y de ahí que se vean empujados hacia adelante, pero la Luna no sólo induce una fuerza sobre la Tierra que a su vez se traduce en una fuerza de inercia sobre los objetos en su superficie, sino que también atrae a los objetos mismos con una fuerza “real” (no inercial), y la combinación de ambas fuerzas tiene un efecto sorprendente.

Para analizarla descomponemos la fuerza total \vec{F} que actúa sobre nuestro objeto en la superficie terrestre como

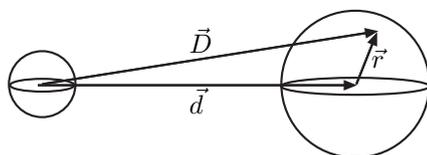
$$\vec{F}_0 + \vec{F}_{\text{gr}} + \vec{F}_{\text{grL}} + \vec{f}_L + \vec{f}_0,$$

donde \vec{F}_{gr} es la atracción gravitatoria terrestre (su peso), \vec{F}_{grL} es la atracción gravitatoria lunar, \vec{F}_0 es el resto de fuerzas reales que actúan sobre el objeto descontadas estas dos, \vec{f}_L es la fuerza de inercia lineal debida a la fuerza que la Luna ejerce sobre la Tierra y \vec{f}_0 es el resto de fuerzas inerciales, incluyendo la fuerza de inercia causada por el Sol. Observemos que

$$\vec{f}_L = \frac{m}{M_T} \vec{F}_L,$$

donde \vec{F}_L es la fuerza con la que la Luna atrae a la Tierra.

Llamamos \vec{d} al vector que une el centro de la Luna con el centro de la Tierra, y \vec{D} al que une el centro de la Luna con la posición del objeto.



Así, según la ley de gravitación universal:

$$\vec{F}_{\text{gr}} = -\frac{GM_T m}{r^3} \vec{r}, \quad \vec{F}_{\text{grL}} = -\frac{GM_L m}{D^3} \vec{D}.$$

Por otra parte,

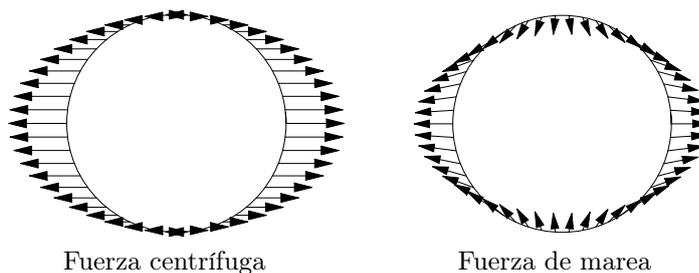
$$\vec{f}_L = \frac{m}{M_T} \frac{GM_T M_L}{d^3} \vec{d} = \frac{GM_L m}{d^3} \vec{d}.$$

Vemos así que la fuerza real \vec{F}_{grL} y la ficticia \vec{f}_L causadas por la Luna sobre el objeto se diferencian en el signo y en que los vectores \vec{D} y \vec{d} no son exactamente el mismo (pero en una figura a escala serían casi idénticos). Por ello conviene estudiar la suma de ambas fuerzas, para tener en cuenta su cancelación parcial:

$$\vec{f}_{\text{mar}} = \vec{F}_{\text{grL}} + \vec{f}_L = -GM_L m \left(\frac{1}{D^3} \vec{D} - \frac{1}{d^3} \vec{d} \right).$$

Esta suma recibe el nombre de *fuerza de marea*, y formalmente es la diferencia entre la fuerza con la que la Luna atrae al objeto menos la fuerza con que lo atraería si éste estuviera en el centro de la Tierra.

La figura muestra la fuerza centrífuga y la fuerza de marea en diferentes puntos de la superficie terrestre, calculadas con las fórmulas que hemos obtenido, aunque están multiplicadas por un factor $k = 10^8$ en el caso de la fuerza centrífuga y $k = 2 \cdot 10^{12}$ en el caso de la fuerza de marea, para que sean apreciables (se supone que la Luna está a la izquierda):



Fuerza centrífuga

Fuerza de marea

La fuerza centrífuga es mucho más intensa que la fuerza de marea. Por ejemplo, en el punto del ecuador más cercano a la Luna, las intensidades son:

$$F_{\text{gr}} = 9.8 \text{ N/kg}, \quad F_{\text{cent}} = 0.3369 \text{ N/kg}, \quad F_{\text{mar}} = 1.1274 \cdot 10^{-6} \text{ N/kg},$$

con lo que la fuerza centrífuga que experimenta un cuerpo es del orden del 0.34% de su peso y la fuerza de marea del orden del 0.000011% de su peso. Son demasiado pequeñas para que puedan apreciarse en un objeto en concreto, pero su efecto se nota en la masa de agua que forma los océanos. La fuerza centrífuga hace que el nivel del mar sea más elevado en el ecuador que en los polos, mientras que la fuerza de marea provoca un efecto similar en los puntos de longitud próxima a la del eje Tierra-Luna.

Sin embargo, una diferencia esencial entre ambas es que la fuerza centrífuga es simétrica respecto del eje de rotación de la Tierra (el eje vertical en la figura), mientras que la fuerza de marea es simétrica respecto del eje Tierra-Luna (el eje horizontal en la figura, de modo que la figura de la derecha es igualmente válida si consideramos que el eje de rotación de la Tierra es el vertical o el perpendicular al papel). Como consecuencia, aunque la fuerza de marea es mucho más débil que la centrífuga, su efecto es más llamativo, ya que, para un observador situado en la Tierra, la Luna gira a su alrededor una vez al día, por lo que los abultamientos que provoca en el nivel del mar giran también una vez al día y, como hay dos abultamientos opuestos, sucede que el nivel del mar sube y baja en cada sitio dos veces al día. Ésta es la explicación del fenómeno de las mareas. El Sol también produce mareas, aunque de menor intensidad, pero las mareas lunares se intensifican sensiblemente cuando la Luna está alineada con el Sol.

Lo sorprendente es que la presencia de la Luna “a la izquierda” de la Tierra provoque una fuerza “hacia la derecha” en la cara opuesta de la Tierra, pero no es tan extraño si nos damos cuenta de que se trata de una fuerza ficticia: cuando un coche acelera “hacia la izquierda”, sus ocupantes sienten una fuerza ficticia “hacia la derecha”. Igualmente, un punto situado en un punto de la superficie terrestre diametralmente opuesto a la Luna, se encuentra con que la Luna “estira” de la Tierra, y la fuerza de inercia correspondiente “empuja” al cuerpo, y la intensidad del empujón es la de la fuerza con que la Luna atraería al objeto si estuviera en el centro de la Tierra, que es mayor que la fuerza con la que atrae al objeto de hecho, porque el centro de la Tierra está más cerca de la Luna. Por ello la fuerza resultante apunta en sentido contrario a la Luna.

El péndulo de Foucault Supongamos que desplazamos un péndulo de su posición de equilibrio y lo dejamos caer sin ninguna velocidad inicial. Teóricamente oscilará siempre en el mismo plano, pero si tiene suficiente masa como para que se haga sensible la fuerza de Coriolis debida a la rotación de la Tierra, ya no será así, sino que el plano de la oscilación sufrirá un movimiento de precesión.

Aunque el hecho ya era conocido, la primera demostración pública de este fenómeno la realizó el físico y astrónomo francés Jean Bernard Léon Foucault, y por ello los péndulos sometidos a la atracción gravitatoria y a la aceleración de Coriolis se llaman *péndulos de Foucault*.

El primer péndulo de Foucault fue instalado en 1851 en la cúpula del *Panteón de París*, estaba sujeto de un cable de acero de 67 m de largo y 1.4 mm de diámetro y pesaba 28 kg. Osciló durante 6 horas con una amplitud máxima

de 6 m y un periodo de 16.5 s. La velocidad de precesión fue de 11° por hora. Vamos a analizar este resultado.

Tomemos un sistema de referencia con origen en el punto donde se sujeta la cuerda del péndulo y supongamos que lo soltamos desde una posición $(0, x_0, z_0)$. Si de momento prescindimos de la fuerza de Coriolis, sobre el péndulo actúan dos fuerzas, su peso $\vec{P} = (0, 0, -mg)$ (en el que podemos considerar incorporada la fuerza centrífuga), y la tensión del cable \vec{T} , en la dirección de éste. Si la posición del péndulo en el instante t es $r(t) = (x(t), y(t), z(t))$, la tensión de la cuerda será

$$\vec{T} = \left(-\frac{T}{l}x, -\frac{T}{l}y, -\frac{T}{l}z \right),$$

donde $l = \sqrt{x^2 + y^2 + z^2}$ es la longitud de la cuerda, que introducimos en la expresión para que $T(t)$ sea precisamente el módulo de la tensión. Por lo tanto, la fuerza total que actúa sobre el péndulo (sin contar la fuerza de Coriolis) es

$$\vec{F} = \left(-\frac{T}{l}x, -\frac{T}{l}y, -\frac{T}{l}z - mg \right).$$

La segunda ley de Newton proporciona el sistema de ecuaciones diferenciales que determinan el movimiento del péndulo:

$$mx''(t) = -\frac{T(t)}{l}x(t), \quad my''(t) = -\frac{T(t)}{l}y(t), \quad mz''(t) = -\frac{T(t)}{l}z(t) - mg.$$

No estamos en condiciones de resolver este sistema, pero pensemos en la función (desconocida) $T(t)$. Podemos calcular su valor en los instantes en los que el péndulo se encontraba a su altura máxima y mínima. En los instantes de altura mínima, es decir, cuando el cable se encontraba en posición vertical, la tensión contrarrestaba exactamente al peso, por lo que $T = 29 \cdot 9.8 = 274.4 \text{ N}$. En los instantes de altura máxima, en los que la velocidad del péndulo era 0, la tensión compensaba la componente normal del peso, con lo que resulta $T = 273.3 \text{ N}$. Éstos son los valores máximo y mínimo de $T(t)$ y, puesto que su diferencia es de 1.1 N , el error que cometemos si suponemos que toma siempre el valor $T = mg = 274.4 \text{ N}$ se mantiene siempre por debajo de un 0.4% .

En general, siempre que consideremos un péndulo con ángulo de oscilación pequeño, obtendremos una aproximación razonable si suponemos que $T \approx mg$, con lo que la tercera ecuación diferencial tiene solución trivial $z(t) = -l$ (lo que indica que estamos despreciando las variaciones en altura del péndulo) y las otras dos se convierten en

$$x''(t) = -\frac{g}{l}x(t), \quad y''(t) = -\frac{g}{l}y(t),$$

que son dos ecuaciones independientes, que podemos resolver por separado.

No perdemos generalidad si suponemos que el péndulo parte de la posición $(0, y_0)$ con velocidad inicial $x'_0 = y'_0 = 0$. (Para asegurarse de que la velocidad inicial era nula, Foucault sujetó el péndulo con un hilo y lo quemó una vez

se hubo estabilizado.) Entonces la primera ecuación diferencial tiene solución trivial $x(t) = 0$, lo que significa que el péndulo oscilará en el plano YZ . Según el teorema 1.35, la segunda ecuación determina una oscilación harmónica, de la forma:

$$y(t) = a \cos\left(\sqrt{\frac{g}{l}}t\right),$$

donde a es la amplitud máxima del péndulo. Es claro entonces que el tiempo que tarda el péndulo en volver a su posición inicial es¹³

$$T = 2\pi\sqrt{l/g}.$$

En el caso del péndulo de París, el valor que obtenemos según nuestros cálculos es $T = 2\pi\sqrt{67/9.8} = 16.43$ s, que encaja muy bien con el valor experimental de 16.5 s.

Hemos llegado de nuevo al resultado que ya habíamos obtenido con otro enfoque al estudiar el péndulo simple, según el cual, para oscilaciones pequeñas, el periodo no depende de la amplitud. Pero el planteamiento que hemos presentado ahora nos permite incorporar fácilmente el efecto de la fuerza de Coriolis, que, según hemos visto, viene dada por $\vec{f}_{\text{cor}} = -2m\vec{\omega} \times \vec{v}_r$.

Recordemos que $\vec{\omega}$ es el vector que apunta en la dirección del eje de rotación de la Tierra y de módulo ω igual a la velocidad angular de rotación de la Tierra. Concretamente,

$$\omega = \frac{2\pi}{24 \cdot 60 \cdot 60} = 7.2722 \cdot 10^{-5} \text{ rad/s}.$$

Podemos suponer que el sistema de coordenadas respecto al cual estamos describiendo el movimiento del péndulo tiene su eje X apuntando hacia el Norte, en cuyo caso el vector $\vec{\omega}$ tiene coordenadas $(\omega \cos \lambda, 0, \omega \sin \lambda)$, donde λ es la latitud. La latitud de París es $\lambda = 48^\circ 50' 46.5''$. Respecto de nuestro sistema de referencia, la fuerza de Coriolis es

$$\begin{aligned} \vec{f}_{\text{cor}} &= -2m\omega(\cos \lambda, 0, \sin \lambda) \times (x', y', z') = \\ &= 2m\omega(y' \sin \lambda, -x' \sin \lambda + z' \cos \lambda, -y' \cos \lambda). \end{aligned}$$

El mayor valor que toma el módulo de y' (en nuestro cálculo sin considerar la fuerza de Coriolis) es $a\sqrt{g/l}$, que en nuestro ejemplo es 2.3 m/s, luego el valor máximo de la componente vertical de la fuerza de Coriolis es

$$2m\omega y'_{\text{max}} \cos \lambda = 0.00615 \text{ N},$$

que es totalmente despreciable en comparación con el peso de 274.4 N. Por otra parte, nuestro supuesto de que la tensión del cable es constante conlleva considerar z constante, con lo que, en coherencia con nuestros supuestos precedentes, debemos considerar $z' = 0$.

¹³Hasta ahora llamábamos T al módulo de la tensión, pero esta cantidad no va a aparecer de nuevo, así que a partir de ahora reservamos la T para referirnos al periodo de oscilación.

En definitiva, una aproximación razonable para la fuerza de Coriolis es

$$\vec{f}_{\text{cor}} = 2m\omega(y' \text{ sen } \lambda, -x' \text{ sen } \lambda, 0),$$

lo que nos lleva a modificar como sigue las ecuaciones del movimiento que habíamos obtenido:

$$x''(t) = 2\omega y'(t) \text{ sen } \lambda - \frac{g}{l}x(t), \quad y''(t) = -2\omega x'(t) \text{ sen } \lambda - \frac{g}{l}y(t).$$

(Notemos que la m desaparece porque las ecuaciones resultan de dividir entre m las dadas por la segunda ley de Newton.) Ahora agrupamos las dos incógnitas en una única función compleja $z(t) = x(t) + y(t)i$, con lo que el sistema de ecuaciones se combina en una única ecuación

$$z''(t) + 2i\omega \text{ sen } \lambda z'(t) + (g/l)z = 0,$$

que se corresponde con la que hemos estudiado en el teorema 1.36. Con la notación empleada allí, $q = \sqrt{g/l + \omega^2 \text{ sen}^2 \lambda}$, luego la solución (correspondiente a una velocidad inicial nula) es

$$z(t) = e^{-i\omega \text{ sen } \lambda t} z_0 \left(\cos qt + i \frac{\omega \text{ sen } \lambda}{q} \text{ sen } qt \right).$$

Para interpretar esta expresión supongamos por un momento que no estuviera la exponencial inicial. La expresión entre paréntesis es una elipse cuyos ejes son los ejes de coordenadas, el semieje mayor mide 1 y el menor $(\omega/q) \text{ sen } \lambda$. La multiplicación por z_0 gira la elipse un ángulo igual al argumento de z_0 , con lo que su semieje mayor pasa a estar en la dirección de z_0 , es decir, del desplazamiento inicial del péndulo, y cuyos semiejes están ahora multiplicados por la amplitud a de este desplazamiento inicial (6 m en el ejemplo que estamos considerando). En total, esta parte de la solución corresponde a una oscilación que ya no es plana, sino elíptica, con un diámetro de 12 metros y el otro de 1.7 mm. En la práctica esto no se distingue de una oscilación plana, con periodo $T = 2\pi/q \approx 2\pi\sqrt{l/g}$ (en nuestro ejemplo $T = 2\pi/q \approx 16.43$ no se distingue del periodo que habíamos calculado sin considerar la fuerza de Coriolis, ya que la corrección es insignificante).

Así pues, si no estuviera la primera exponencial, la solución que hemos obtenido sería prácticamente la misma que la que ya teníamos, pero dicho factor se interpreta como un giro de la elipse (es decir, del plano de oscilación) con velocidad angular $\omega_\lambda = \omega \text{ sen } \lambda$ (en sentido horario en el hemisferio Norte y antihorario en el hemisferio Sur). En nuestro ejemplo, la velocidad angular es

$$\omega_\lambda = 5.47558 \cdot 10^{-5} \text{ rad/s} = 11^\circ 17' 39.1''/\text{h},$$

que cuadra con el resultado experimental. ■

Sistemas binarios Hemos visto que, a la hora de elegir un sistema de referencia para describir el movimiento de unos cuerpos, lo importante no es si es o no inercial, sino tener en cuenta las fuerzas de inercia en caso de que las haya y no podamos considerarlas despreciables. Por ejemplo, al estudiar el movimiento de un planeta bajo la atracción del Sol hemos tomado un sistema de referencia con origen en el Sol, pero no hemos tenido en cuenta que dicho sistema de referencia no es inercial, ya que el Sol está sometido a la atracción gravitatoria del planeta, luego desde un sistema de referencia inercial el centro del Sol está sometido a una aceleración. Lo que sucede es que dicha aceleración es despreciable debido a la gran diferencia de masas entre el Sol y el planeta (es como la aceleración que experimenta la Tierra por la atracción de una piedra que cae sobre ella).

Ahora bien, si consideramos, por ejemplo, el movimiento de un sistema binario de estrellas de masas similares sometidas a su propia atracción gravitatoria, ya no podemos tomar un sistema de referencia centrado en una de ellas y despreciar la fuerza de inercia que afectará a la otra. Por lo tanto, en principio, toda la discusión de las leyes de Kepler que hemos expuesto no es aplicable a este caso. No obstante, vamos a ver que en realidad sí que lo es.

Supongamos que tenemos dos objetos de masas m_1 y m_2 situados en posiciones \vec{r}_1 y \vec{r}_2 respecto de un sistema de referencia inercial, sometidos únicamente a las fuerzas gravitatorias que se ejercen mutuamente. Si llamamos $\vec{r} = \vec{r}_2 - \vec{r}_1$, éstas son

$$\vec{F}_{12} = -\frac{Gm_1m_2}{r^3}\vec{r}, \quad \vec{F}_{21} = -\vec{F}_{12} = \frac{Gm_1m_2}{r^3}\vec{r}.$$

Ahora consideramos el *centro de masas* del sistema formado por los dos cuerpos, que es el punto dado por

$$\vec{R} = \frac{m_1\vec{r}_1 + m_2\vec{r}_2}{2}.$$

Es un punto del espacio perfectamente identificado por los dos cuerpos, aunque no haya nada en él. Si los dos cuerpos tienen masas iguales, el centro de masas es el punto medio entre ambos, pero si no, el centro de masas estará más cerca del más masivo, de forma proporcional. Por ejemplo, si $m_1 = 2m_2$, entonces \vec{R} está al doble de distancia del segundo cuerpo que del primero.

El interés de este punto es que, derivando dos veces, la aceleración del centro de masas es

$$\vec{A} = \frac{m_1\vec{a}_1 + m_2\vec{a}_2}{2} = \frac{\vec{F}_{21} + \vec{F}_{12}}{2} = \vec{0},$$

donde suponemos que \vec{F}_{21} es la única fuerza que se ejerce sobre el primer cuerpo y \vec{F}_{12} es la única fuerza que se ejerce sobre el segundo, por lo que $\vec{F}_{21} = m_1\vec{a}_1$, $\vec{F}_{12} = m_2\vec{a}_2$. Esto significa que, aunque los dos objetos estén sometidos a fuerzas no nulas y ello haga que estén acelerados respecto de un cualquier de referencia inercial, su centro de masas no lo está.¹⁴

¹⁴Esto se generaliza fácilmente a sistemas de cualquier cantidad finita de partículas puntuales: si sobre ellas no actúan fuerzas exteriores, es decir, distintas de las fuerzas que cada partícula ejerce sobre las demás, y que se compensan a pares según la tercera ley de Newton, el centro de masas tiene aceleración nula respecto de cualquier sistema de referencia inercial.

Por lo tanto, podemos tomar un sistema de referencia inercial con origen de coordenadas en el centro de masas de los dos cuerpos. Esto se traduce en la relación adicional $m_1\vec{r}_1 + m_2\vec{r}_2 = \vec{0}$, de donde

$$\vec{r} = \vec{r}_2 - \vec{r}_1 = \frac{m_1 + m_2}{m_1} \vec{r}_2, \quad r = \frac{m_1 + m_2}{m_1} r_2.$$

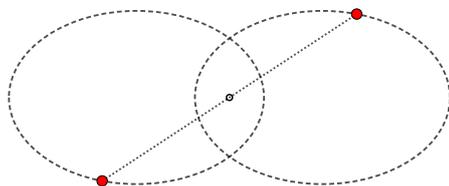
Por lo tanto,

$$\vec{F}_{12} = -\frac{Gm_1m_2}{r^3} \vec{r} = -\frac{Gm_1^*m_2^*}{r_2^3} \vec{r}_2, \quad \text{donde } m_i^* = \frac{m_1}{m_1 + m_2} m_i.$$

Esto significa que, respecto del sistema inercial con origen en el centro de masas, la fuerza que actúa sobre el segundo cuerpo está dirigida hacia el origen de coordenadas y su intensidad es la que daría la ley de gravitación universal si las masas de los cuerpos fueran m_i^* , por lo que el cuerpo de masa m_2 se moverá siguiendo las leyes de Kepler, con las únicas salvedades de que en el foco de su órbita no estará el objeto de masa m_1 , sino el centro de masas del sistema.

Naturalmente, todo vale igualmente si intercambiamos los papeles, luego concluimos que dos objetos que se mueven por la fuerza gravitatoria que se ejercen mutuamente, no orbitan uno alrededor del otro, sino que ambos orbitan (según las leyes de Kepler) alrededor de su centro de masas, y esto no depende del sistema de referencia (inercial) que se considere.

Por ejemplo, si tenemos dos estrellas con la misma masa, ambas girarán en elipses respectivas con un foco en común en su punto medio:



Pero si m_1 es mucho mayor que m_2 , entonces $m_i^* \approx m_i$ y el centro de masas estará muy próximo a \vec{r}_1 , por lo que en la práctica podemos decir que el segundo cuerpo gira alrededor del primero, como ocurre con una estrella y un planeta.

■

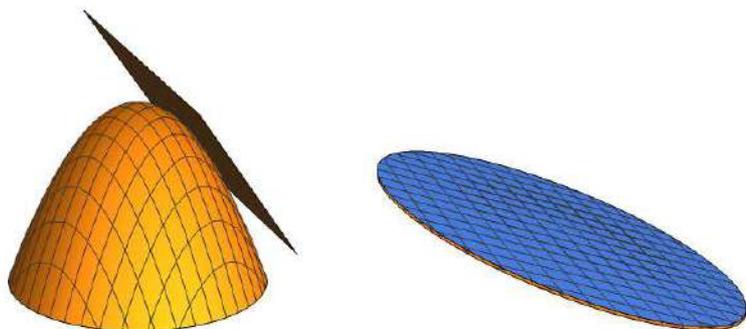
Capítulo V

Diferenciabilidad

Hemos visto que una función $f : A \subset \mathbb{R} \rightarrow \mathbb{R}$ definida en un abierto A es derivable en un punto $c \in A$ si la gráfica de f se vuelve indistinguible de una recta (no vertical) cuando se amplía alrededor de c , y dicha recta es la que llamamos recta tangente a la gráfica de f en el punto considerado.

Ahora vamos a estudiar la situación para una función $f : A \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, es decir, vamos a ver bajo qué condiciones la gráfica de f tiene un plano tangente en uno de sus puntos $(c, f(c))$, un plano que pase por dicho punto y que, cuando se amplía suficientemente la gráfica de f , resulte indistinguible de ella.

Por ejemplo, la figura de la izquierda es la gráfica de $f(x, y) = 10 - x^2 - y^2$ junto con su plano tangente en el punto $c = (1, 2)$. A la derecha vemos las mismas gráficas restringidas a los puntos (x, y) del círculo de centro $(1, 2)$ y de radio 0.2, y vemos que resultan casi indistinguibles. Si consideráramos únicamente los puntos del círculo de radio 0.1, la diferencia sería ya inapreciable:



Un plano que pase por un punto $(x_0, y_0, f(x_0, y_0))$ (a menos que sea vertical) es la gráfica de una función de la forma

$$T(x, y) = f(x_0, y_0) + a(x - x_0) + b(y - y_0).$$

La pregunta obligada es cuánto tienen que valer concretamente las constantes a y b para que el plano $T(x, y)$ tenga la propiedad de tangencia que hemos descrito.

5.1 Funciones diferenciables

Recordemos que las aplicaciones $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ de la forma

$$\phi(x, y) = ax + by$$

son las llamadas aplicaciones lineales. En estos términos, la discusión precedente puede formularse como que buscamos una aplicación lineal ϕ con la propiedad de que el plano dado por la gráfica de la función

$$T(x, y) = f(x_0, y_0) + \phi(x - x_0, y - y_0)$$

se vuelva indistinguible de $f(x, y)$ cuando ambas gráficas se amplían alrededor del punto $c = (x_0, y_0)$. Si llamamos $v = (x - x_0, y - y_0)$, lo queremos es que

$$f(c + v) \approx f(c) + \phi(v)$$

en el sentido de que ambos miembros resulten indistinguibles cuando se amplían las gráficas. Ahora bien, si ampliamos las gráficas de modo que la longitud $\|v\|$ pase a verse como de una unidad, todas las longitudes se dividen entre $\|v\|$, luego la diferencia entre ambos miembros pasa a verse de longitud

$$\frac{f(c + v) - f(c) - \phi(v)}{\|v\|}.$$

Es esta cantidad la que queremos que se vuelva arbitrariamente pequeña a medida que la escala $1/\|v\|$ de la ampliación es mayor, es decir, a medida que v tiende a 0. Por consiguiente, la aplicación lineal ϕ que corresponde al plano tangente debe cumplir que

$$\lim_{v \rightarrow 0} \frac{f(c + v) - f(c) - \phi(v)}{\|v\|} = 0.$$

Esto nos lleva a la definición siguiente:

Definición 5.1 Diremos que una aplicación $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ definida en un abierto A es *diferenciable* en un punto $c \in A$ si existe una aplicación lineal $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$, es decir, de la forma $\phi(x) = a_1x_1 + \cdots + a_nx_n$, de modo que

$$\lim_{v \rightarrow 0} \frac{f(c + v) - f(c) - \phi(v)}{\|v\|} = 0.$$

En tal caso, la gráfica de la función $T(x) = f(c) + \phi(x - c)$ se llama *variedad¹ tangente* a la gráfica de f en el punto c .

Antes de analizar el contenido de esta definición necesitamos hacer algunas observaciones técnicas:

¹La palabra “variedad” se usa aquí como término genérico para incluir los casos “recta” (si $n = 1$) “plano” (si $n = 2$) o “espacio” (si $n = 3$).

- Hemos dado la definición para funciones con dominio en \mathbb{R}^n para incluir los casos $n = 1, 2, 3$, que son los que realmente nos van a interesar.
- En la sección 1.1 hemos definido el concepto de “abierto” para subconjuntos $A \subset \mathbb{K}$, lo cual incluye los casos $A \subset \mathbb{R}$ y $A \subset \mathbb{C} = \mathbb{R}^2$, pero la definición vale igualmente para \mathbb{R}^3 y también para \mathbb{R}^n si definimos $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$. La repetimos en general:

Un conjunto $A \subset \mathbb{R}^n$ es *abierto* si para todo punto $c \in A$ existe un $\eta > 0$ tal que todo $x \in A$ que cumpla $\|x - c\| < \eta$ está también en A .

- Similarmente, en [ITAn 3.12] definimos el concepto de límite para funciones $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$, lo cual incluye los casos $A \subset \mathbb{R}$ y $A \subset \mathbb{C} = \mathbb{R}^2$. Nuevamente, la definición vale en general, aunque sólo necesitamos el caso particular que aparece en la definición de diferenciabilidad que acabamos de dar:

Sea $B = \{x \in \mathbb{R}^n \mid \|x - c\| < \eta\}$, sea $g : B \setminus \{c\} \rightarrow \mathbb{R}$ y sea $l \in \mathbb{R}$. Diremos que $\lim_{x \rightarrow c} g(x) = l$ si para todo $\epsilon > 0$ existe un $\delta > 0$ ($\delta < \eta$) tal que si $0 < \|x - c\| < \delta$ entonces $|f(x) - l| < \epsilon$.

- Por último, aunque no ha aparecido en la definición de diferenciabilidad, conviene generalizar también el concepto de continuidad de una función en un punto:

Diremos que una función $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es *continua* en un punto $c \in A$ si para todo $\epsilon > 0$ existe un $\delta > 0$ tal que si $x \in A$ cumple $\|x - c\| < \delta$ entonces $|f(x) - f(c)| < \epsilon$.

Pasemos a analizar ya la definición de diferenciabilidad para el caso $n = 2$. Supongamos que hemos fijado una unidad de medida “apreciable”, como 1 cm, y fijemos un $\epsilon > 0$ que sea menor que nuestro umbral de discernimiento visual, por ejemplo, $\epsilon = 0.0001$ (una micra).

Si ϕ cumple la definición de diferenciabilidad de f en el punto c , por la definición de abierto existe un $\eta_0 > 0$ tal que si $x \in A$ y $\|x - c\| < \eta_0$, entonces $x \in A$ o, equivalentemente, si $v \in \mathbb{R}^n$ y $\|v\| < \eta_0$, entonces $x + v \in A$.

Por la definición de límite existe $0 < \eta < \eta_0$ tal que si $0 < \|v\| < \eta$, entonces

$$\left| \frac{f(c+v) - f(c) - \phi(v)}{\|v\|} \right| < \frac{\epsilon}{5}.$$

Fijemos ahora $\eta < \delta$ y vamos a considerar la ampliación de la gráfica de f en el círculo $\|c - x\| < \eta$ que hace que este círculo de radio η pase a tener radio 5 cm, lo que equivale a tomar como factor de escala $5/\eta$. Cada punto del círculo puede expresarse como $x = c + v$, donde $\|v\| \leq \eta < \delta$. Entonces, la

diferencia $|f(x) - T(x)|$ entre la función f y el plano T pasará a ser en la figura a escala

$$5 \left| \frac{f(c+v) - f(c) - \phi(v)}{\|v\|} \right| < 5 \frac{\epsilon}{5} < \epsilon,$$

luego en efecto, la representación a escala de la gráfica de f sobre todo el círculo de centro c y radio η estará sobre un círculo de 10 cm de diámetro en el que la diferencia entre $f(x)$ y $T(x)$ será siempre inferior a una micra.

Si queremos una precisión aún mayor sólo tendremos que tomar un δ menor, pero siempre podremos encontrar un círculo de centro c que, al ser ampliado hasta que tenga 10 cm de diámetro, la gráfica de f se diferencie de la del plano tangente en menos del margen ϵ prefijado. Obviamente el haber fijado la unidad de medida en 1 cm o el haber exigido que la base de la gráfica ampliada mida concretamente 10 cm de diámetro es irrelevante, pues el razonamiento vale igualmente con otros datos cualesquiera.

En conclusión, la definición de diferenciabilidad (para $n = 2$) afirma precisamente que el plano

$$T(x) = f(c) + \phi(x - c)$$

es tangente a la gráfica de f en el sentido que hemos discutido.

Antes de analizar los casos $n = 1$ o $n = 3$, conviene probar algunos hechos generales sobre la diferenciabilidad de funciones. Observemos en primer lugar que no sabríamos cómo justificar si una función dada f es diferenciable o no en un punto c de su dominio porque para ello tendríamos que considerar una aplicación lineal ϕ en concreto, y no sabemos cuál es la adecuada para que se cumpla la definición. Vamos a ocuparnos en primer lugar de este asunto.

Para ello observamos que si ϕ cumple la definición de diferenciabilidad, para todo $v \in \mathbb{R}^n$ no nulo se cumple que

$$\lim_{h \rightarrow 0} \frac{f(c + hv) - f(c) - \phi(hv)}{\|hv\|} = 0,$$

donde este límite es el límite de una función de una variable $h \in \mathbb{R}$, que hemos estudiado en [ITAn]. En efecto, dado $\epsilon > 0$, por la definición de límite (en \mathbb{R}^n) existe un $\eta > 0$ tal que si $0 < \|w\| < \eta$ entonces

$$\left| \frac{f(c+w) - f(c) - \phi(w)}{\|w\|} \right| < \epsilon,$$

y a su vez, tomando $\delta = \eta/\|v\|$, tenemos que si $|h| < \delta$ entonces $\|hv\| < \eta$, luego

$$\left| \frac{f(c+hv) - f(c) - \phi(hv)}{\|hv\|} \right| < \epsilon$$

y esto prueba que el límite existe y vale 0.

Ahora usamos que $\phi(hv) = h\phi(v)$ y que $\|hv\| = |h|\|v\|$ (junto con las propiedades que conocemos para límites de funciones de una variable). Vemos que

$$\frac{1}{\|v\|} \lim_{h \rightarrow 0} \frac{f(c+hv) - f(c) - h\phi(v)}{|h|} = 0.$$

Eliminamos la norma y usamos que la función $|h|/h$ está acotada en $\mathbb{R} \setminus \{0\}$, así como que el producto de una función acotada por otra que tiende a 0 también tiende a 0, luego

$$\lim_{h \rightarrow 0} \frac{f(c + hv) - f(c) - h\phi(v)}{h} = 0,$$

pero esto equivale a

$$\lim_{h \rightarrow 0} \frac{f(c + hv) - f(c)}{h} - \phi(v) = 0,$$

o también a que

$$\phi(v) = \lim_{h \rightarrow 0} \frac{f(c + hv) - f(c)}{h}.$$

Definición 5.2 Dada una función $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ definida en un abierto $A \subset \mathbb{R}^n$, llamaremos *derivada direccional* de f en el punto $c \in A$ y en la dirección $v \in \mathbb{R}^n$, $v \neq 0$ al límite

$$f'(c; v) = \lim_{h \rightarrow 0} \frac{f(c + hv) - f(c)}{h}.$$

Hemos demostrado que si f es diferenciable en c , la aplicación lineal ϕ que cumple la definición es única, pues en cada $v \in \mathbb{R}^n$ no nulo tiene que venir dada por

$$\phi(v) = f'(c; v).$$

(Además, notemos que toda aplicación lineal cumple $\phi(0) = 0$.)

Por lo tanto, si f es diferenciable en c , podemos definir la *diferencial* de f en c como la única aplicación lineal $df(c) : \mathbb{R}^n \rightarrow \mathbb{R}$ que cumple la definición de diferenciable, es decir, que cumple

$$\lim_{v \rightarrow 0} \frac{f(c + v) - f(c) - df(c)(v)}{\|v\|} = 0.$$

Hemos demostrado el teorema siguiente:

Teorema 5.3 Si $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es una función diferenciable en el punto $c \in A$, entonces tiene derivadas direccionales en todas las direcciones $v \in \mathbb{R}^n$ no nulas, y además vienen dadas por

$$f'(c; v) = df(c)(v).$$

Las derivadas direccionales son fáciles de calcular y tienen una interpretación muy simple. Basta observar que si definimos $g(h) = f(c + hv)$, entonces

$$g'(0) = \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} = \lim_{h \rightarrow 0} \frac{f(c + hv) - f(c)}{h} = f'(c; v).$$

Ejemplo Vamos a calcular las derivadas direccionales de $f(x, y) = 10 - x^2 - y^2$ en el punto $c = (1, 2)$. Para ello, si $v = (p, q)$, consideramos la función

$$g(h) = f((1, 2) + h(p, q)) = f(1 + hp, 2 + hq) = 10 - (1 + hp)^2 - (2 + hq)^2,$$

con lo que

$$g'(h) = -2(1 + hp)p - 2(2 + hq)q,$$

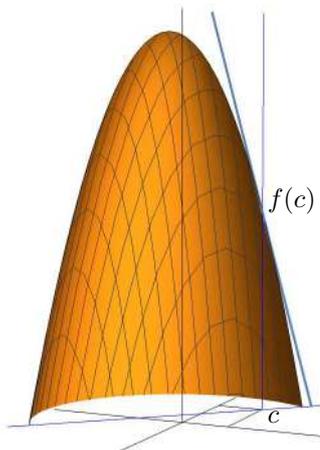
y concluimos que

$$f'((1, 2); (p, q)) = g'(0) = -2p - 4q.$$

No hemos probado que f sea diferenciable en $(1, 2)$, pero ahora sabemos que, si lo es, su diferencial tiene que ser $df(1, 2)(p, q) = -2p - 4q$, luego el plano tangente en $(1, 2)$ tiene que ser

$$T(x, y) = f(1, 2) - 2(x - 1) - 4(y - 2) = 5 - 2(x - 1) - 4(y - 2).$$

En cuanto a la interpretación geométrica de la derivada direccional, ésta es más simple si consideramos vectores de norma 1, como $v = (\sqrt{2}/2, \sqrt{2}/2)$. La figura muestra la gráfica de f cortada por el plano vertical situado sobre la recta $y = x + 1$, que pasa por $c = (1, 2)$ con vector director v (o también $(1, 1)$).



Observemos que $c + hv$ es el punto que está a h unidades de distancia de c sobre la recta que pasa por c con vector director v (en la dirección de v si $h > 0$ y en la contraria si $h < 0$, por lo que la función auxiliar $g(h) = f(c + hv)$ es la que tiene por gráfica el borde del corte, si tomamos como eje de abscisas la recta $y = x + 1$ con origen en c y como eje de ordenadas el eje vertical. La derivada $g'(0)$ es, por consiguiente, la pendiente de la tangente a dicho borde en $(0, g(0))$, es decir, en el punto $(c, f(c))$, que en este caso es

$$f'(1, 2)\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) = -3\sqrt{2}.$$

Esto significa que por cada unidad que avanzamos a partir de c en la dirección de v , la función f disminuye aproximadamente $-3\sqrt{2}$ unidades. Si el vector v no es unitario la diferencia es que la derivada direccional no indica la variación de f por cada unidad de avance, sino por cada paso de longitud $\|v\|$ que avanzamos.

Más en general, para cualquier n , se cumple que si una función f es derivable en la dirección de v en un punto c , también lo es en cualquier dirección αv , con $\alpha \neq 0$, y $f'(c; \alpha v) = \alpha f'(c; v)$. En efecto:

$$f'(c; \alpha v) = \lim_{h \rightarrow 0} \frac{f(c + h\alpha v) - f(c)}{h} = \alpha \lim_{h \rightarrow 0} \frac{f(c + h\alpha v) - f(c)}{h\alpha} = \alpha f'(c; v).$$

Por ejemplo, si $\|v\| = 2$, entonces $f'(c; v)$ es el doble de la pendiente de la gráfica de f en la dirección de v , pues por cada paso que demos de longitud $\|v\|$ habremos avanzado 2 unidades, y la tangente habrá variado el doble de su pendiente.

Así pues, la derivada direccional nos da la pendiente de la gráfica de f en la dirección de v , medida en unidades de variación de f por número de pasos de longitud $\|v\|$ dados. ■

Ahora observamos que si una función es diferenciable y

$$df(c)(x) = a_1 x_1 + \cdots + a_n x_n,$$

entonces

$$a_i = df(c)(0, \dots, \overset{(i)}{1}, \dots, 0) = f'(c; e_i),$$

donde $e_i = (0, \dots, \overset{(i)}{1}, \dots, 0)$. Esto nos lleva a otra definición:

Definición 5.4 Si $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es una función definida en un abierto A , su derivada parcial respecto de la variable x_i en un punto $c \in A$ es la derivada direccional

$$\left. \frac{\partial f}{\partial x_i} \right|_c = f'(c; e_i) = \lim_{h \rightarrow 0} \frac{f(c_1, \dots, c_i + h, \dots, c_n) - f(c)}{h}.$$

Hemos demostrado que si f es diferenciable en c , entonces tiene derivadas direccionales en c , luego en particular tiene derivadas parciales, y

$$df(c)(v) = \left. \frac{\partial f}{\partial x_1} \right|_c v_1 + \cdots + \left. \frac{\partial f}{\partial x_n} \right|_c v_n.$$

Si una función f tiene derivada parcial i -ésima en todos los puntos de su dominio, llamamos

$$\frac{\partial f}{\partial x_i} : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$$

a la función que a cada punto c le asigna la derivada parcial de f en c .

Ahora observamos que las derivadas parciales son muy fáciles de calcular, sin más que tener en cuenta que, en las condiciones de la definición anterior, se cumple que

$$\left. \frac{\partial f}{\partial x_i} \right|_c = g'(c_i),$$

donde $g(x) = f(c_1, \dots, c_{i-1}, x, c_{i+1}, \dots, c_n)$. En efecto, si escribimos la definición de $g'(c_i)$ obtenemos exactamente la definición de derivada parcial de f . Así pues:

La derivada parcial i -ésima de una función f en un punto x es la derivada de la función de una variable que resulta de considerar constantes todas las variables x_j excepto x_i .

Y esto a su vez nos da su interpretación:

La derivada parcial i -ésima de una función f en un punto x indica la tasa de variación instantánea de f por cada unidad que aumenta la variable x_i supuesto que las demás variables no se modifiquen.

Esto hay que entenderlo con todos los matices que señalamos al discutir la interpretación de las derivadas de funciones de una variable.

Así pues, la existencia y el cálculo de derivadas parciales se justifican trivialmente a partir de los resultados que conocemos para funciones de una variable.

Ejemplo Consideremos la fórmula

$$T(l, g) = 2\pi\sqrt{\frac{l}{g}}$$

que calcula el periodo de oscilación de un péndulo en función de la longitud l de la cuerda y la intensidad g del campo gravitatorio.

Podemos afirmar que

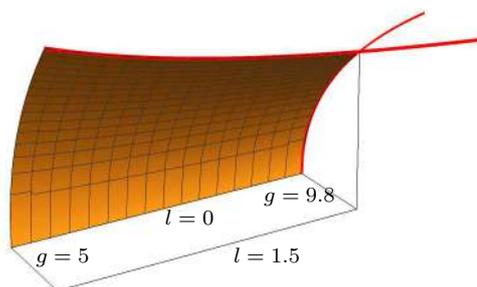
$$\frac{\partial T}{\partial l} = \frac{\pi}{\sqrt{lg}}, \quad \frac{\partial T}{\partial g} = -\pi\sqrt{\frac{l}{g^3}}.$$

En efecto, si consideramos a T como función de l , para un valor fijo de g , los resultados de la sección 1.2 sobre cálculo de derivadas nos aseguran que se trata de una función derivable y que su derivada tiene el valor que hemos indicado, e igualmente sucede con la derivada respecto de g .

Por ejemplo,

$$\left. \frac{\partial T}{\partial l} \right|_{(1.5, 9.8)} = 0.819, \quad \left. \frac{\partial T}{\partial g} \right|_{(1.5, 9.8)} = -0.125,$$

Y esto significa que si tenemos un péndulo de 1.5 m de longitud y $g = 9.8$ N/kg, por cada metro que alarguemos la cuerda, su periodo de oscilación aumentará 0.819 segundos. Esto, literalmente, no es cierto, pues el periodo es de $T(1.5, 9.8) = 2.46$ s y, si ampliamos la cuerda, pasa a ser de $T(2.5, 9.8) = 3.17$ s, con lo que el aumento ha sido de 0.715 s, con lo que el error es de casi un 15%, pero es que la derivada tiene que entenderse como una tasa de variación instantánea. Por ejemplo, para un incremento de 1 cm en la longitud del péndulo, el periodo aumenta en $T(1.51, 9.8) - T(1.5, 9.8) = 0.00818$ s, y la derivada predice $0.819 \cdot 0.01 = 0.00819$, y ahora el error es de un 0.12%.



Geométicamente, las derivadas parciales son las derivadas de las funciones de una variable cuyas gráficas son las curvas que resultan de cortar la gráfica de T con los planos verticales situados sobre las rectas $g = 9.8$ y $l = 1.5$, respectivamente, es decir, las pendientes de las tangentes a dichas curvas en el punto $(1.5, 9.8, T(1.5, 9.8))$. ■

Extremos locales La definición 1.9 de máximos y mínimos locales de funciones de una variable se generalizan trivialmente a funciones de varias variables:

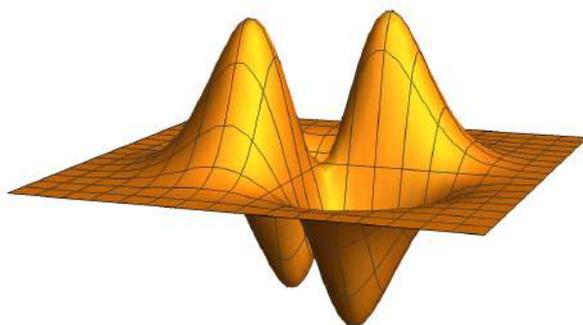
Dada una función $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$, se dice que un punto c es un *máximo local* (resp. *mínimo local*) de f si existe un $\delta > 0$ tal que para todo $x \in A$ tal que $\|x - c\| < \delta$ se cumple que $f(c) \geq f(x)$ (resp. $f(c) \leq f(x)$).

Si A es abierto y f es derivable en A , una condición necesaria para que un punto c pueda ser un máximo o un mínimo local de f es que sus derivadas parciales se anulen en c , pues si $\partial f / \partial x_i|_c > 0$ (resp. < 0), esto significa que un pequeño aumento de la variable x_i a partir de c hace aumentar (resp. disminuir) la función, y una disminución de x_i hace disminuir (resp. aumentar) la función, luego ésta no puede tener ni un máximo local ni un mínimo local en c .

Ejemplo Consideremos la función

$$f(x, y) = 4xy e^{1-x^2-y^2},$$

cuya gráfica se muestra en la figura:



En ella se ve que tiene dos máximos y dos mínimos locales (que de hecho son globales, es decir, son puntos en los que la función alcanza su valor máximo y su valor mínimo). Puesto que es obviamente derivable, una forma de localizarlos es igualar a 0 las derivadas parciales:

$$\frac{\partial f}{\partial x} = 4ye^{1-x^2-y^2} + 4xye^{1-x^2-y^2}(-2x) = 0,$$

$$\frac{\partial f}{\partial y} = 4xe^{1-x^2-y^2} + 4xye^{1-x^2-y^2}(-2y) = 0,$$

lo cual equivale a

$$y - 2x^2y = 0, \quad x - 2xy^2 = 0,$$

y este sistema de ecuaciones tiene claramente cinco soluciones:

$$(0, 0), \quad \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \quad \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), \quad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \quad \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right),$$

en las cuales f toma los valores $0, 2, 2, -2, -2$, respectivamente, luego el primer punto, pese a anular las derivadas, no es ni un máximo ni un mínimo local (es fácil ver que la función toma valores positivos y negativos tan cerca de $(0, 0)$ como se quiera) y los siguientes son los dos máximos y los dos mínimos. ■

Ejemplo: Las funciones coordenadas Consideremos la función

$$\pi_i : \mathbb{R}^n \longrightarrow \mathbb{R}$$

dada por $\pi_i(x) = x_i$, es decir, la función que a cada n -tupla le hace corresponder su coordenada i -ésima. Igual que podemos hablar de la función $x^4y + y^2$, es costumbre referirse a la función π_i simplemente como la función x_i . Claramente es derivable, y sus derivadas parciales valen

$$\frac{\partial x_i}{\partial x_j} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Por lo tanto, por la observación tras la definición 5.4, si x_i es diferenciable, su diferencial tiene que ser

$$dx_i(c)(v) = v_i.$$

Pero es fácil ver que, ciertamente, x_i es diferenciable en cualquier punto c :

$$\lim_{v \rightarrow 0} \frac{x_i(c+v) - x_i(c) - v_i}{\|v\|} = \lim_{v \rightarrow 0} \frac{c_i + v_i - c_i - v_i}{\|v\|} = \lim_{v \rightarrow 0} \frac{0}{\|v\|} = \lim_{v \rightarrow 0} 0 = 0.$$

Como consecuencia, si $f : A \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ es cualquier función diferenciable en todo punto de $c \in A$, tenemos que

$$df(c)(v) = \frac{\partial f}{\partial x_1} \Big|_c v_1 + \cdots + \frac{\partial f}{\partial x_n} \Big|_c v_n = \frac{\partial f}{\partial x_1} \Big|_c dx_1(c)(v) + \cdots + \frac{\partial f}{\partial x_n} \Big|_c dx_n(c)(v).$$

Como esta igualdad es cierta para todo vector v , tenemos la igualdad de funciones

$$df(c) = \frac{\partial f}{\partial x_1} \Big|_c dx_1(c) + \cdots + \frac{\partial f}{\partial x_n} \Big|_c dx_n(c).$$

A su vez, como esto es cierto para todo punto c , podemos expresar esta relación en la forma

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n. \quad \blacksquare$$

Ejemplo Vamos a probar que la función $f(x, y) = 10 - x^2 - y^2$ es diferenciable en \mathbb{R}^2 . Sabemos que, si lo es, su diferencial tiene que ser necesariamente

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = -2x dx - 2y dy,$$

así que vamos a comprobar que se cumple la definición de diferenciabilidad con esta aplicación lineal en concreto:

$$\begin{aligned} \lim_{(u,v) \rightarrow (0,0)} \frac{f(x+u, y+v) - f(x, y) + 2xu + 2yv}{\sqrt{u^2 + v^2}} &= \\ \lim_{(u,v) \rightarrow (0,0)} \frac{10 - x^2 - 2xu - u^2 - y^2 - v^2 - 2yv - 10 + x^2 + y^2 + 2xu + 2yv}{\sqrt{u^2 + v^2}} &= \\ \lim_{(u,v) \rightarrow (0,0)} \frac{-u^2 - v^2}{\sqrt{u^2 + v^2}} &= \lim_{(u,v) \rightarrow (0,0)} -\sqrt{u^2 + v^2} = 0. \end{aligned}$$

En la sección siguiente veremos que en la práctica todas las funciones “usuales” (en particular, los polinomios como $10 - x^2 - y^2$) son diferenciables en su dominio, de modo que los cálculos como el que acabamos de hacer resultan innecesarios. ■

5.2 Ejemplos y resultados fundamentales

En la sección anterior hemos introducido los conceptos necesarios para definir la diferenciabilidad y para dar sentido a expresión

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n.$$

para la diferencial de una función diferenciable en un punto arbitrario de su dominio. Ahora vamos a discutir las propiedades fundamentales de las funciones diferenciables y a ilustrarlas con ejemplos oportunos. Por propiedades “fundamentales” nos referimos a algunos resultados que, aunque su enunciado es muy simple y natural, varios de ellos tienen pruebas técnicas que pueden resultar un tanto artificiosas, así que, para evitar que estos tecnicismos desvíen la atención del lector respecto de las ideas de fondo y de los ejemplos, en estos casos enunciaremos los teoremas para su discusión y pospondremos las demostraciones hasta la sección siguiente. Una vez contemos con estos hechos básicos estaremos en condiciones de trabajar “limpiamente” con las funciones diferenciables.

Empezamos observando que la diferenciabilidad de funciones de una variable no aporta nada nuevo:

Teorema 5.5 *Una función $f : A \subset \mathbb{R} \rightarrow \mathbb{R}$ es diferenciable en un punto $c \in A$ si y sólo si es derivable en c .*

DEMOSTRACIÓN: Sabemos que si f es diferenciable en c , entonces tiene derivada parcial respecto de x en c , pero en este caso la definición de derivada parcial coincide con la definición de derivada:

$$\left. \frac{\partial f}{\partial x} \right|_c = \lim_{h \rightarrow 0} \frac{f(c+h) - f(c)}{h} = f'(c).$$

Recíprocamente, supongamos que f es derivable en c . Entonces tenemos que existe el límite anterior, luego

$$\lim_{h \rightarrow 0} \frac{f(c+h) - f(c) - f'(c)h}{h} = 0.$$

Ahora usamos que la función $h/|h|$ está acotada en $\mathbb{R} \setminus \{0\}$, y que el producto de una función acotada por otra que tiende a 0 tiende también a 0, luego multiplicando el límite anterior por $h/|h|$ queda que

$$\lim_{h \rightarrow 0} \frac{f(c+h) - f(c) - f'(c)h}{|h|} = 0,$$

pero $|h| = \|h\|$, luego esto es justo la definición de diferenciabilidad de f en c . ■

En las condiciones del teorema anterior, la recta tangente definida en 5.1 no es sino

$$T(x) = f(c) + f'(c)(x - c),$$

que es la recta tangente definida en 1.1, y la diferencial

$$df = \frac{df}{dx} dx$$

es la misma definida en 1.38.

Consideremos ahora el caso de funciones de dos variables. Ahora en cada punto tenemos (o podemos tener) dos derivadas parciales e infinitas derivadas direccionales. Si la función es diferenciable, la diferencial está determinada por las derivadas parciales, pero que éstas existan no garantiza que la función sea diferenciable.

Ejemplo Consideremos la función

$$f(x, y) = \begin{cases} 0 & \text{si } xy = 0, \\ 1 & \text{si } xy \neq 0. \end{cases}$$

Esta función tiene derivadas parciales en $(0, 0)$, pero no es diferenciable. En efecto,

$$\frac{\partial f}{\partial x} \Big|_{(0,0)} = \lim_{h \rightarrow 0} \frac{f(h, 0) - f(0, 0)}{h} = \lim_{h \rightarrow 0} \frac{0}{h} = \lim_{h \rightarrow 0} 0 = 0.$$

Igualmente se concluye que la derivada respecto de y vale 0. Sin embargo, si la función fuera diferenciable, también existiría la derivada direccional

$$f'((0, 0); (u, v)) = \lim_{h \rightarrow 0} \frac{f(hu, hv) - f(0, 0)}{h} = \lim_{h \rightarrow 0} \frac{1}{h} = \infty$$

siempre que $uv \neq 0$.

No tiene nada de extraño que una función pueda tener derivadas parciales en un punto, es decir, que al cortar su gráfica por los planos verticales paralelos a los ejes de coordenadas se obtengan curvas derivables, pero que al cortar la gráfica con cualquier otro plano vertical se obtenga una función no derivable. ■

En realidad la función del ejemplo anterior no es diferenciable por un hecho más obvio que la no existencia de derivadas direccionales, y es que claramente es una función discontinua en $(0, 0)$. Sin embargo:

Teorema 5.6 *Toda función diferenciable en un punto es continua en él.*

La demostración de este teorema no es muy complicada, pero por coherencia la posponemos hasta la sección siguiente (página 316) junto con los enunciados siguientes más técnicos cuya prueba vamos a posponer.

Ejemplo Consideremos la función

$$f(x, y) = \begin{cases} \frac{x^3 - y^3}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$

Vamos a ver que es continua en $(0, 0)$ y que tiene derivadas direccionales en todas las direcciones, pero a pesar de ello no es diferenciable. En efecto, f es continua en $(0, 0)$, pues si $(x, y) = (\rho \cos \theta, \rho \sin \theta)$ es la expresión en coordenadas polares de un punto no nulo, tenemos que

$$f(x, y) = \rho \frac{\cos^3 \theta - \sin^3 \theta}{\cos^2 \theta + \sin^2 \theta} = \rho(\cos^3 \theta - \sin^3 \theta),$$

luego, teniendo en cuenta que $\rho = \|(x, y)\|$, tenemos que

$$|f(x, y)| \leq \|(x, y)\|(|\cos \theta|^3 + |\sin \theta|^3) \leq 2\|(x, y)\|,$$

luego si $\|(x, y)\| < \epsilon/2$ se cumple que $|f(x, y) - f(0, 0)| < \epsilon$, lo que prueba la continuidad.

Además, existen todas las derivadas direccionales en $(0, 0)$:

$$f'((0, 0); (u, v)) = \lim_{h \rightarrow 0} \frac{f(hu, hv) - f(0, 0)}{h} = \lim_{h \rightarrow 0} \frac{u^3 - v^3}{u^2 + v^2} = \frac{u^3 - v^3}{u^2 + v^2}.$$

En particular, tomando $(u, v) = (1, 0), (0, 1)$,

$$\left. \frac{\partial f}{\partial x} \right|_{(0,0)} = 1, \quad \left. \frac{\partial f}{\partial y} \right|_{(0,0)} = -1.$$

Por lo tanto, si f fuera diferenciable, tendría que ser

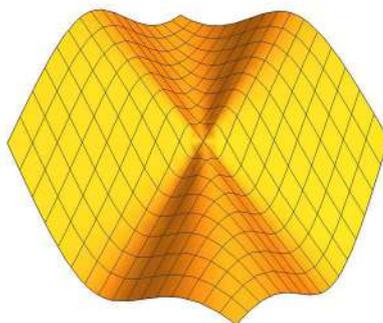
$$f'((0, 0); (2, 1)) = \left. \frac{\partial f}{\partial x} \right|_{(0,0)} \cdot 2 + \left. \frac{\partial f}{\partial y} \right|_{(0,0)} \cdot 1 = 1,$$

cuando hemos visto que

$$f'((0, 0); (2, 1)) = \frac{2^3 - 1^3}{2^2 - 1^2} = \frac{7}{3}.$$

Por lo tanto, f no es diferenciable.

La figura siguiente muestra lo que sucede: la función es continua en $(0, 0)$, y tiene rectas tangentes en todas las direcciones, pero éstas no se disponen sobre un mismo plano, por lo que no hay un plano tangente. Si aumentamos la gráfica alrededor del punto $(0, 0)$, nunca dejaremos de verla “arrugada”, por lo que nunca la confundiremos con un plano:



Pero todavía podemos poner un ejemplo más sutil:

Ejemplo Consideremos la función

$$f(x, y) = \begin{cases} \frac{x^5 y^5}{x^{12} + y^8} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$

Vamos a probar que es continua en $(0, 0)$, que tiene rectas tangentes en todas las direcciones y que éstas forman el plano $z = 0$, pero aun así no es diferenciable.

Para probar la continuidad observamos que, por la desigualdad entre la media aritmética y la geométrica,

$$\frac{x^{12} + y^8}{13} = \frac{\overbrace{\frac{x^{12}}{5} + \cdots + \frac{x^{12}}{5}}^5 + \overbrace{\frac{y^8}{8} + \cdots + \frac{y^8}{8}}^8}{13} \geq \frac{\sqrt[13]{x^{60} y^{64}}}{5^5 8^8},$$

luego, llamando $C = 5^5 8^8 / 13$,

$$C(x^{12} + y^8)|x|^{5/13}|y|^{1/13} \geq |x|^5|y|^5,$$

luego

$$\left| \frac{x^5 y^5}{x^{12} + y^8} \right| \leq C \sqrt[13]{|x|^5|y|^5},$$

y ahora la continuidad es inmediata. Por otra parte,

$$f'((0,0);(u,v)) = \lim_{h \rightarrow 0} \frac{f(hu, hv) - f(0,0)}{h} = \lim_{h \rightarrow 0} h \frac{u^5 v^5}{h^3 u^{12} + v^8} = 0,$$

luego todas las derivadas direccionales de f en $(0,0)$ valen 0. A su vez, esto significa que la gráfica de f tiene rectas tangentes en $(0,0)$ en todas las direcciones, y que todas ellas forman el plano $z = 0$. Sin embargo, f no es diferenciable en $(0,0)$. La diferenciable equivale a que

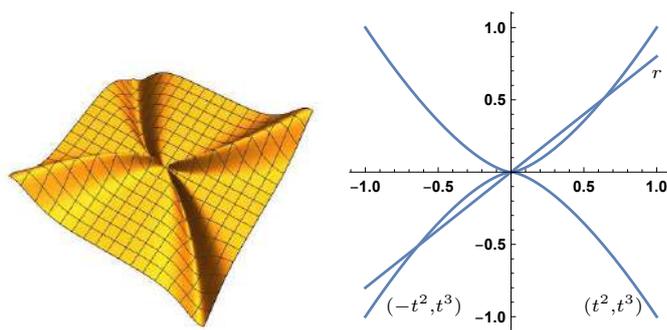
$$\lim_{(u,v) \rightarrow (0,0)} \frac{f(u,v)}{\sqrt{u^2 + v^2}} = 0,$$

pero esto no es cierto, porque si consideramos, en particular, puntos de la forma $(u,v) = (t^2, t^3)$, vemos que

$$\left| \frac{f(u,v)}{\sqrt{u^2 + v^2}} \right| = \frac{|t|^{25}/(t^{24} + t^{24})}{\sqrt{t^4 + t^6}} = \frac{1}{2|t|\sqrt{1+t^2}},$$

luego vemos que hay puntos arbitrariamente próximos a $(0,0)$ donde esta expresión es arbitrariamente grande.

La figura siguiente muestra la gráfica de f alrededor de $(0,0)$. Vemos que es bastante plana salvo alrededor de las curvas (t^2, t^3) y $(-t^2, t^3)$, sobre cada una de las cuales tiene una cresta y un valle. Y por mucho que ampliemos la gráfica alrededor de $(0,0)$ nunca dejaremos de ver esas crestas y esos valles que delatan que la superficie no es plana.



Sobre cualquier recta r que pasa por $(0,0)$, en un entorno de $(0,0)$, la gráfica tiende a parecerse a la propia recta r (es decir, tiene pendiente 0), y por eso las

derivadas direccionales son 0, pero en cuanto la gráfica sobre r se aleja de 0 para acercarse a las curvas $(\pm t^2, t^3)$, crece o decrece para formar la cresta o el valle correspondiente. Como cualquier entorno de $(0, 0)$ contiene parte de las crestas y los valles, la función no es diferenciable en $(0, 0)$, pues las crestas y los valles distinguen a su gráfica del plano $z = 0$ en cualquier entorno de $(0, 0)$. ■

Aunque esto pueda sugerir que es difícil reconocer cuándo una función es diferenciable, sucede que hay una condición suficiente muy simple que basta para probar la diferenciabilidad de cualquier función “razonable”:

Teorema 5.7 Sea $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ una función definida en un abierto A . Si tiene derivadas parciales en A y éstas son continuas, entonces f es diferenciable en todos los puntos de A .

La prueba es bastante técnica, así que la posponemos hasta la sección siguiente (página 316).

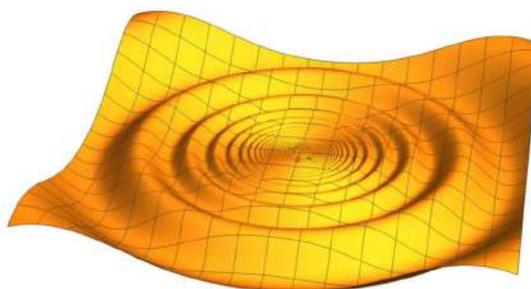
Ejemplo Ahora podemos afirmar que la función $f(x, y) = 10 - x^2 - y^2$ es diferenciable en \mathbb{R}^2 simplemente porque tiene derivadas parciales

$$\frac{\partial f}{\partial x} = -2x, \quad \frac{\partial f}{\partial y} = -2y,$$

que son continuas porque son polinomios, lo que hace innecesario el cálculo con el que habíamos probado esto mismo al final de la sección anterior. ■

Ejemplo Consideremos la función

$$f(x, y) = \begin{cases} (x^2 + y^2) \operatorname{sen} \frac{1}{\sqrt{x^2 + y^2}} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$



Vamos a probar que es diferenciable en $(0, 0)$, pero sus derivadas no son continuas, por lo que la condición suficiente del teorema anterior no es necesaria.

Para probar la diferenciabilidad observamos que

$$\lim_{(u,v) \rightarrow (0,0)} \frac{f(u, v) - f(0, 0)}{\sqrt{u^2 + v^2}} = \lim_{(u,v) \rightarrow (0,0)} \sqrt{u^2 + v^2} \operatorname{sen} \frac{1}{\sqrt{x^2 + y^2}} = 0,$$

pues es el producto de una función que tiende a 0 por otra acotada. Por otra parte:

$$\left. \frac{\partial f}{\partial x} \right|_{(0,0)} = \lim_{h \rightarrow 0} h \operatorname{sen} \frac{1}{|h|} = 0,$$

por el mismo motivo, e igualmente se anula la derivada respecto de y . Por lo tanto, el plano tangente a la gráfica de la función en $(0, 0)$ es $z = 0$.

Por otra parte, en un punto (x, y) distinto de $(0, 0)$, podemos usar las reglas usuales de derivación para calcular

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2x \operatorname{sen} \frac{1}{\sqrt{x^2 + y^2}} + (x^2 + y^2) \cos \frac{1}{\sqrt{x^2 + y^2}} \left(-\frac{1}{2}\right) (x^2 + y^2)^{-3/2} 2x \\ &= 2x \operatorname{sen} \frac{1}{\sqrt{x^2 + y^2}} - \frac{x}{\sqrt{x^2 + y^2}} \cos \frac{1}{\sqrt{x^2 + y^2}}, \end{aligned}$$

y no existe

$$\lim_{(x,y) \rightarrow (0,0)} \frac{\partial f}{\partial x},$$

pues, si $h > 0$

$$\left. \frac{\partial f}{\partial x} \right|_{(h,h)} = 2h \operatorname{sen} \frac{1}{h\sqrt{2}} - \frac{1}{\sqrt{2}} \cos \frac{1}{h\sqrt{2}},$$

y el primer sumando tiende a 0 cuando $h \rightarrow 0$, pero el segundo oscila entre $\pm 1/\sqrt{2}$, luego no converge.

En la representación gráfica vemos que oscila una y otra vez alrededor de $(0, 0)$, pero las oscilaciones son cada vez más pequeñas, de modo que si ampliáramos mucho la gráfica, lo que veríamos no se distinguiría de un plano porque las oscilaciones serían microscópicas. Así pues, tenemos un ejemplo de función diferenciable con derivadas parciales discontinuas, lo que prueba que la condición suficiente del teorema anterior no es necesaria. ■

Veamos un ejemplo de aplicación del teorema 5.7:

Teorema 5.8 Sea $f :]a, b[\times]c, d[\rightarrow \mathbb{R}$ una función continua tal que exista

$$\frac{\partial f(t, z)}{\partial z} :]a, b[\times]c, d[\rightarrow \mathbb{R}$$

y sea continua. Entonces la función $g :]a, b[\times]a, b[\times]c, d[\rightarrow \mathbb{R}$ dada por

$$g(x, y, z) = \int_x^y f(t, z) dt$$

tiene derivadas parciales continuas, luego es diferenciable. Más concretamente:

$$\frac{\partial g}{\partial x} = -f(x, z), \quad \frac{\partial g}{\partial y} = f(y, z), \quad \frac{\partial g}{\partial z} = \int_x^y \left. \frac{\partial f}{\partial z} \right|_{(t,z)} dt.$$

DEMOSTRACIÓN: Fijemos un punto (x_0, y_0, z_0) en el dominio de g y vamos a demostrar que las derivadas parciales son continuas en dicho punto. Podemos tomar $a < a' < x_0, y_0 < b' < b, c < c' < z_0 < d' < d$, de modo que $\partial f/\partial z$ es continua en el compacto $[a', b'] \times [c', d']$.

Si $F(t)$ es una primitiva de la función continua $f(t, z_0)$, entonces

$$g(x, y, z_0) = F(y) - F(x),$$

de donde

$$\begin{aligned} \left. \frac{\partial g}{\partial x} \right|_{x_0, y_0, z_0} &= -F'(x_0) = -f(x_0, z_0), \\ \left. \frac{\partial g}{\partial y} \right|_{x_0, y_0, z_0} &= F'(y_0) = f(y_0, z_0). \end{aligned}$$

En particular es inmediato que estas dos derivadas parciales son continuas. La expresión de la derivada respecto de z viene dada por el teorema 3.11, teniendo en cuenta la nota posterior. Para probar su continuidad observamos que

$$\begin{aligned} \left| \left. \frac{\partial g}{\partial z} \right|_{(x, y, z)} - \left. \frac{\partial g}{\partial z} \right|_{(x_0, y_0, z_0)} \right| &= \left| \int_x^y \left. \frac{\partial f}{\partial z} \right|_{(t, z)} dt - \int_{x_0}^{y_0} \left. \frac{\partial f}{\partial z} \right|_{(t, z_0)} dt \right| \leq \\ &\int_x^{x_0} \left| \left. \frac{\partial f}{\partial z} \right|_{(t, z)} \right| dt + \int_{y_0}^y \left| \left. \frac{\partial f}{\partial z} \right|_{(t, z)} \right| dt + \int_{x_0}^{y_0} \left| \left. \frac{\partial f}{\partial z} \right|_{(t, z)} - \left. \frac{\partial f}{\partial z} \right|_{(t, z_0)} \right| dt. \end{aligned}$$

Sea M una cota de $\partial f/\partial z$ en el compacto $[a', b'] \times [c', d']$ y, por el teorema de Heine-Cantor [ITAn A.14], dado $\epsilon > 0$, podemos tomar un $\delta > 0$ tal que si $\|(t, z) - (t', z')\| < \delta$, entonces

$$\left| \left. \frac{\partial f}{\partial z} \right|_{(t, z)} - \left. \frac{\partial f}{\partial z} \right|_{(t, z_0)} \right| < \frac{\epsilon}{3(b-a)}.$$

Podemos exigir que $\delta < \epsilon/3M$. Así, si $\|(x, y, z) - (x_0, y_0, z_0)\| < \delta$, tenemos que

$$\left| \left. \frac{\partial g}{\partial z} \right|_{(x, y, z)} - \left. \frac{\partial g}{\partial z} \right|_{(x_0, y_0, z_0)} \right| \leq M|x - x_0| + M|y - y_0| + \frac{\epsilon}{3(b-a)}|y_0 - x_0| < \epsilon.$$

■

Derivadas sucesivas Como en el caso de funciones de una variable, si una derivada parcial de una función tiene a su vez derivadas parciales, podemos volverla a derivar, y lo que obtenemos es una derivada segunda de la función dada. Si ésta tiene a su vez derivadas parciales, éstas son las derivadas terceras de la función dada, y así sucesivamente. La notación habitual para las derivadas sucesivas de una función es:

$$\frac{\partial^7 f}{\partial x^2 \partial y \partial x \partial y^3}.$$

Esto hay que entenderlo como la función que resulta de derivar sucesivamente 7 veces la función f , la primera vez respecto de x , luego otra vez respecto de x , luego respecto de y , luego respecto de x y luego otras tres veces respecto de y . Se dice que es una de las derivadas de orden 7 de la función f .

Definición 5.9 Se dice que una función $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ definida en un conjunto abierto A es de clase C^k en A , donde $k \geq 1$ es un número natural, si admite todas las derivadas parciales de orden k y todas ellas son continuas en A . Se dice que f es de clase C^∞ en A si admite derivadas de todos los órdenes y todas ellas son continuas en A . En este contexto conviene llamar funciones de clase C^0 en A a las funciones continuas en A .

En estos términos el teorema 5.7 afirma que toda función de clase C^1 en un abierto es diferenciable en él, y lo que sucede es que prácticamente todas las funciones diferenciables “de interés” son, de hecho, de clase C^∞ , y en muchos casos lo son trivialmente, sin más que tener en cuenta el teorema siguiente y las observaciones posteriores:

Teorema 5.10 *Se cumple:*

- a) Las funciones coordenadas $x_i : \mathbb{R}^n \rightarrow \mathbb{R}$ son de clase C^∞ .
- b) Si $f : A \subset \mathbb{R} \rightarrow \mathbb{R}$ está definida por una serie de potencias, entonces es de clase C^∞ .
- c) Si $f, g : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ son funciones de clase C^∞ en su dominio A , también lo son $f + g$, αf (para cualquier $\alpha \in \mathbb{R}$), fg , si g no se anula, también f/g .
- d) Si $f : A \subset \mathbb{R}^n \rightarrow B \subset \mathbb{R}$ y $g : B \subset \mathbb{R} \rightarrow \mathbb{R}$ son funciones de clase C^∞ , también lo es la composición $g(f(x))$.

DEMOSTRACIÓN: a) Es inmediato, pues las derivadas de primer orden de x_i son constantes (valen 0 o 1), luego las derivadas de segundo orden son nulas, al igual que las derivadas de orden n para todo $n \geq 2$, luego todas ellas son continuas.

Los apartados siguientes se prueban fácilmente por inducción teniendo en cuenta que una función es de clase C^{k+1} si y sólo si existen todas sus derivadas de orden 1 y todas ellas son de clase C^k (pues las derivadas de orden $k+1$ son las derivadas de orden k de las derivadas de primer orden).

b) Toda serie de potencias define una función continua, luego es de clase C^0 . Supuesto que las series de potencias sean de clase C^k , el teorema 1.22 afirma que la derivada de f es también una serie de potencias, luego por hipótesis de inducción f' es de clase C^k , luego f es de clase C^{k+1} . Como esto vale para todo k , de hecho f es de clase C^∞ .

c) Todos los casos se prueban igual. Veamos por ejemplo el caso del cociente. Sabemos² que un cociente de funciones de clase C^0 (donde el denominador no se anula) es una función de clase C^0 . Supongamos ahora que el cociente de funciones de clase C^k es de clase C^k y que f y g son de clase C^{k+1} . Sabemos que el cociente de funciones derivables (de una variable) es derivable, y la derivada viene dada por

$$\frac{\partial(f/g)}{\partial x_i} = \frac{\partial f/\partial x_i g - f \partial g/\partial x_i}{g^2}.$$

Como $\partial f/\partial x_i$ es de clase C^k y g es de clase C^{k+1} (luego también es de clase C^k), por la propiedad para productos, tenemos que $\partial f/\partial x_i g$ es de clase C^k e igualmente se razona con el segundo término del numerador, luego (por la propiedad de la suma) éste es de clase C^k . De nuevo por la propiedad para productos, g^2 es de clase C^k , luego por hipótesis de inducción la derivada es de clase C^k , y esto prueba que f/g es de clase C^{k+1} .

d) Sabemos que el teorema es cierto para funciones de clase C^0 . Si es cierto para funciones de clase C^k y las funciones f y g son de clase C^{k+1} , entonces, por la regla de la cadena, la función $h(x) = g(f(x))$ (considerada sólo como función de x_i , con las demás coordenadas fijas) es derivable, y su derivada vale

$$\frac{\partial h}{\partial x_i} = \frac{\partial g}{\partial t} \Big|_{f(x)} \frac{\partial f}{\partial x_i}.$$

La derivada $\partial g/\partial t$ es de clase C^k , luego por hipótesis de inducción, al componerla con la función f , que es de clase C^{k+1} , luego en particular de clase C^k , obtenemos una función de clase C^k , y al multiplicarla por $\partial f/\partial x_i$, que también es de clase C^k , por el apartado anterior obtenemos una función de clase C^k , luego $\partial h/\partial x_i$ es de clase C^k , luego h es de clase C^{k+1} . Como esto vale para todo k , concluimos que h es de clase C^∞ . ■

En particular:

- Todo polinomio de n variables es de clase C^∞ , incluso todo cociente de polinomios restringido a un abierto donde el denominador no se anule (por los apartados a) y c))
- Por a) son C^∞ las funciones e^x , $\sen x$, $\cos x$, $\sinh x$, $\cosh x$. Por c) también lo son $\tan x$ y $\tanh x$ donde están definidas, por ser cocientes de funciones C^∞ .
- La función $1/x$ es C^∞ en $\mathbb{R} \setminus \{0\}$, luego $\log x$ es C^∞ en $]0, +\infty[$ porque su derivada es C^∞ .

²En realidad en [ITAn] tenemos probada la versión de este teorema para funciones continuas de una o dos variables. Para probar (los apartados c y d de) este teorema para $n \geq 3$ necesitamos que los mismos argumentos empleados allí para funciones de dos variables valen igualmente para cualquier número de variables, o al menos para 3 variables, que es el único caso que vamos a necesitar. Para un análisis más detallado véanse las observaciones en la página 318.

- La función $\sqrt[n]{x} = e^{(1/n)\log x}$ es C^∞ en $]0, +\infty[$ por el apartado d). De hecho si n es impar lo es en $\mathbb{R} \setminus \{0\}$ pues $\sqrt[n]{x} = -\sqrt[n]{-x}$.
- Si f y g son C^∞ y $g > 0$, entonces $g^f = e^{f \log g}$ es C^∞ por c) y d).
- Las funciones $\arcsen x$, $\arccos x$, $\arctan x$ son C^∞ porque los casos anteriores prueban que sus derivadas lo son.

Con esto tenemos asegurado que cualquier función “usual” es de clase C^∞ .

Otro hecho fundamental sobre derivadas sucesivas es que no dependen del orden de derivación:

Teorema 5.11 (Teorema de Schwarz) *Si $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es una función de clase C^2 en A , entonces*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Para la demostración véase la página 318.

Ejemplo Consideremos la función $f(x, y) = x^y = e^{y \log x}$, definida para $x > 0$. Sus derivadas de primer orden son

$$\frac{\partial f}{\partial x} = yx^{y-1}, \quad \frac{\partial f}{\partial y} = x^y \log x.$$

Si las derivamos respecto a la variable opuesta obtenemos:

$$\frac{\partial^2 f}{\partial x \partial y} = x^{y-1} + yx^{y-1} \log x, \quad \frac{\partial^2 f}{\partial y \partial x} = yx^{y-1} \log x + x^y (1/x).$$

No ha quedado exactamente lo mismo, pero operando $x^y(1/x) = x^{y-1}$ comprobamos que se cumple el teorema de Schwarz. Notemos que si no supiéramos que yx^{y-1} y $x^y \log x$ son las derivadas parciales de una misma función, no habría motivos para sospechar que la derivada de la primera respecto de y tenga que coincidir con la derivada de la segunda respecto de x . ■

Notemos que el teorema de Schwarz justifica cualquier intercambio del orden de derivación en cualquier derivada sucesiva (siempre que las derivadas sean continuas), pues, por ejemplo,

$$\frac{\partial^4 f}{\partial x \partial z \partial y \partial x} = \frac{\partial^4 f}{\partial x \partial z \partial x \partial y} = \frac{\partial^4 f}{\partial x^2 \partial z \partial y} = \frac{\partial^4 f}{\partial x^2 \partial y \partial z}.$$

La primera igualdad resulta de aplicar el teorema de Schwarz a la función

$$\frac{\partial^2 f}{\partial x \partial z},$$

luego aplicamos el teorema a la función $\partial f / \partial x$, que nos da

$$\frac{\partial^3 f}{\partial x \partial z \partial x} = \frac{\partial^3 f}{\partial x^2 \partial z}$$

y derivando respecto de y obtenemos la segunda igualdad. Por último aplicamos el teorema de Schwarz a la función

$$\frac{\partial^2 f}{\partial x^2}$$

y obtenemos la última igualdad.

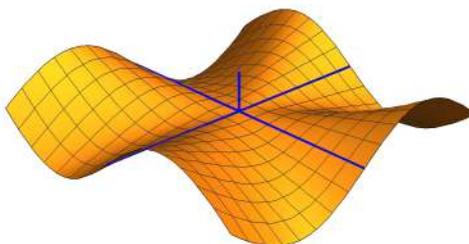
En particular, toda derivada sucesiva de una función se puede expresar en la forma

$$\frac{\partial^k f}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}},$$

con las variables en orden.

Ejemplo Consideremos la función

$$f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{si } (x, y) = (0, 0). \end{cases}$$



A la vista de su gráfica todo apunta a que es diferenciable en \mathbb{R}^2 . De hecho, es una función de clase C^1 . Para probarlo observamos en primer lugar que f se anula sobre los ejes, lo que implica inmediatamente que

$$\left. \frac{\partial f}{\partial x} \right|_{(0,0)} = \left. \frac{\partial f}{\partial y} \right|_{(0,0)} = 0.$$

En los demás puntos, un cálculo rutinario muestra que

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{2x^2y}{x^2 + y^2} - \frac{2x^2y(x^2 - y^2)}{(x^2 + y^2)^2} + \frac{y(x^2 - y^2)}{x^2 + y^2}, \\ \frac{\partial f}{\partial y} &= -\frac{2xy^2}{x^2 + y^2} - \frac{2xy^2(x^2 - y^2)}{(x^2 + y^2)^2} + \frac{x(x^2 - y^2)}{x^2 + y^2}, \end{aligned}$$

y se cumple que

$$\lim_{(x,y) \rightarrow (0,0)} \frac{\partial f}{\partial x} = \lim_{(x,y) \rightarrow (0,0)} \frac{\partial f}{\partial y} = 0.$$

Para probarlo sustituimos $x = \rho \cos \theta$, $y = \rho \operatorname{sen} \theta$, con lo que

$$\frac{\partial f}{\partial x} = \rho 2 \cos^2 \theta \operatorname{sen} \theta - \rho 2 \cos^2 \theta \operatorname{sen} \theta \cos 2\theta + \rho \cos 2\theta,$$

luego

$$\left| \frac{\partial f}{\partial x} \right| \leq 5 \|(x, y)\|,$$

y lo mismo sucede con la derivada respecto de y , luego las derivadas parciales son continuas en \mathbb{R}^2 . Sin embargo,

$$\frac{\partial f}{\partial x} \Big|_{(0,y)} = -y, \quad \frac{\partial f}{\partial y} \Big|_{(0,y)} = -x,$$

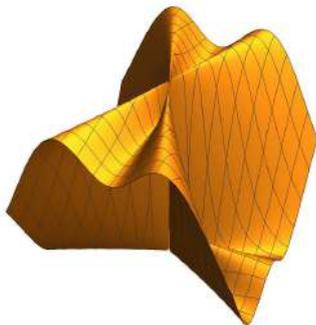
de donde

$$\frac{\partial^2 f}{\partial x \partial y} \Big|_{(0,0)} = -1, \quad \frac{\partial^2 f}{\partial y \partial x} \Big|_{(0,0)} = 1.$$

El teorema de Schwarz implica que f no es de clase C^2 . No obstante, claramente es de clase C^∞ en $\mathbb{R}^2 \setminus \{(0,0)\}$, luego, por el teorema de Schwarz, se cumple que

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

en todos los puntos distintos de $(0,0)$. La figura siguiente muestra la gráfica de esta función:



Sobre el eje x vale siempre 1 y sobre el eje y vale siempre -1 , pero en $(0,0)$ vale 1 o -1 según si se calcula derivando primero respecto de x o respecto de y . ■

La regla de la cadena Discutimos ahora la versión de la regla de la cadena para funciones diferenciables. Vamos a ilustrarlo primero con un ejemplo:

Consideremos una función cualquiera, como $z(x, y) = 3x^2 + 5y^2$. Claramente es de clase C^∞ y su diferencial es

$$dz = 6x dx + 10y dy. \quad (5.1)$$

Podemos expresar z en términos de las coordenadas polares de los puntos, es decir, sustituyendo

$$x = \rho \cos \theta, \quad y = \rho \operatorname{sen} \theta. \quad (5.2)$$

Así obtenemos otra función que es natural llamar

$$z(\rho, \theta) = \rho^2(3 \cos^2 \theta + 5 \operatorname{sen}^2 \theta).$$

También es diferenciable y

$$\begin{aligned} dz &= 2\rho(3 \cos^2 \theta + 5 \operatorname{sen}^2 \theta) d\rho + \rho^2(6 \cos \theta \operatorname{sen} \theta - 10 \operatorname{sen} \theta \cos \theta) d\theta \\ &= \rho(6 \cos^2 \theta + 10 \operatorname{sen}^2 \theta) d\rho - 2\rho^2 \operatorname{sen} 2\theta d\theta. \end{aligned} \quad (5.3)$$

Por otro lado, las funciones que definen el cambio de coordenadas también son diferenciables, y

$$dx = \cos \theta d\rho - \rho \operatorname{sen} \theta d\theta, \quad dy = \operatorname{sen} \theta d\rho + \rho \cos \theta d\theta. \quad (5.4)$$

La “magia” aparece cuando sustituimos (5.2) y (5.4) en (5.1) obtenemos

$$dz = 6\rho \cos \theta (\cos \theta d\rho - \rho \operatorname{sen} \theta d\theta) + 10\rho \operatorname{sen} \theta (\operatorname{sen} \theta d\rho + \rho \cos \theta d\theta),$$

y si operamos llegamos exactamente a (5.3). Si al lector le parece “lógico”, debería pararse a pensar que en (5.1) las funciones dx y dy son las diferenciales de las proyecciones $(x, y) \mapsto x$, $(x, y) \mapsto y$, mientras que en (5.4) son otra cosa, pues son las diferenciales de las funciones que definen el cambio a polares. Similarmente, dz en (5.1) es la diferencial de una función distinta $z(x, y)$ a la función $z(\rho, \theta)$ cuya diferencial es (5.3). Precisamente el resultado parece “lógico” porque la notación está elegida para que la sustitución de (5.2) y (5.4) en (5.1) parezca una inocente manipulación algebraica, pero lo cierto es que no lo es, sino que es el resultado analítico que vamos a discutir a continuación y que, en una primera aproximación, podríamos enunciar así:

Si una magnitud z está en función de otras x, y y éstas a su vez están en función de otras $x(\rho, \theta), y(\rho, \theta)$, entonces, la diferencial $dz(\rho, \theta)$ de la función $z(\rho, \theta)$ que resulta de sustituir x, y por sus expresiones en función de ρ, θ se puede calcular sustituyendo en $dz(x, y)(dx, dy)$ las variables x, y por sus expresiones en función de ρ, θ y las diferenciales dx, dy por sus expresiones en función de $\rho, \theta, d\rho, d\theta$.

El enunciado general es el siguiente:

Teorema 5.12 (Regla de la cadena) *Sea $z(y_1, \dots, y_n)$ una función diferenciable (resp. de clase C^∞) en un abierto $A \subset \mathbb{R}^n$ y sean $y_i = y_i(x_1, \dots, x_m)$ funciones diferenciables (resp. de clase C^∞) definidas en un mismo dominio $y_i : B \subset \mathbb{R}^m \rightarrow \mathbb{R}$ de modo que determinen una función $y : B \subset \mathbb{R}^m \rightarrow A$, que a su vez define una función compuesta $z(x_1, \dots, x_m)$ dada por*

$$z(x_1, \dots, x_m) = z(y_1(x_1, \dots, x_m), \dots, y_n(x_1, \dots, x_m)).$$

Entonces la función compuesta es diferenciable (resp. de clase C^∞) en B y, para cada punto $x \in B$, se cumple que

$$dz(x)(v) = dz(y_1(x), \dots, y_n(x))(dy_1(x)(v), \dots, dy_n(x)(v)).$$

La regla de la cadena admite una formulación más explícita y práctica:

Tenemos que

$$dz = \sum_{i=1}^n \frac{\partial z}{\partial y_i} dy_i$$

y a su vez

$$dy_i = \sum_{j=1}^m \frac{\partial y_i}{\partial x_j} dx_j,$$

luego el teorema afirma que la diferencial de la función $z(x)$ es

$$dz = \sum_{j=1}^m \frac{\partial z}{\partial x_j} dx_j = \sum_{i=1}^n \frac{\partial z}{\partial y_i} \bigg|_{y(x)} \sum_{j=1}^m \frac{\partial y_i}{\partial x_j} dx_j = \sum_{j=1}^m \left(\sum_{i=1}^n \frac{\partial z}{\partial y_i} \bigg|_{y(x)} \frac{\partial y_i}{\partial x_j} \right) dx_j.$$

Por consiguiente, evaluando en el vector $e_j = (0, \dots, 1, \dots, 0)$ (con el 1 en la posición j -ésima), obtenemos que

$$\frac{\partial z}{\partial x_j} = \sum_{i=1}^n \frac{\partial z}{\partial y_i} \bigg|_{y(x)} \frac{\partial y_i}{\partial x_j}. \quad (5.5)$$

Es esta fórmula analítica (y no una manipulación algebraica) lo que explica por qué al sustituir (5.2) y (5.4) en (5.1) se obtiene (5.3) en el ejemplo anterior, y ahora tal vez el lector concluya que es “lógico” que así sea, pero por una consideración analítica: lo que expresa esta fórmula es que la variación de z debida a una variación unitaria de x_j puede calcularse multiplicando la variación de z debida a una variación unitaria de y_i por la variación de y_i debida a una variación unitaria de x_j y sumando todas estas variaciones.

De todos modos, por “lógica” que pueda parecer, la prueba de la regla de la cadena no es trivial y la veremos en la página 319. Como aplicación demostramos la fórmula (3.3), que habíamos dejado pendiente:

Teorema 5.13 (Regla de Leibniz) *Si $f :]a^*, b^*[\times]c, d[\rightarrow \mathbb{R}$ es una función continua con derivada respecto de x también continua y $a, b :]u, v[\rightarrow]a^*, b^*[$ son funciones derivables, entonces*

$$\frac{d}{dx} \int_{a(x)}^{b(x)} f(t, x) dt = f(b(x), x) b'(x) - f(a(x), x) a'(x) + \int_{a(x)}^{b(x)} \frac{\partial f}{\partial x} dt.$$

DEMOSTRACIÓN: Basta aplicar la regla de la cadena. Por 5.8 sabemos que la función

$$g(u, v, x) = \int_u^v f(t, x) dt$$

es de clase C^1 y la función que queremos derivar es $h(x) = g(a(x), b(x), x)$, luego la igualdad (5.5) se particulariza en este caso a

$$h'(x) = \frac{\partial g}{\partial u} \bigg|_{(a(x), b(x), x)} a'(x) + \frac{\partial g}{\partial v} \bigg|_{(a(x), b(x), x)} b'(x) + \frac{\partial g}{\partial x} \bigg|_{(a(x), b(x), x)}.$$

Teniendo en cuenta que las derivadas de g son las calculadas en 5.8, obtenemos la fórmula del enunciado. ■

Funciones vectoriales En el apartado anterior hemos considerado el cambio a coordenadas polares, pero es más natural verlo como una aplicación $P :]0, +\infty[\times \mathbb{R} \rightarrow \mathbb{R}^2$, dada por

$$P(\rho, \theta) = (\rho \cos \theta, \rho \operatorname{sen} \theta).$$

En general, una función $f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$ está determinada por dos funciones coordenadas, $f_1, f_2 : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$.

Definición 5.14 Diremos que una función $f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$ es *diferenciable* en un punto c si lo son sus funciones coordenadas $f_1, f_2 : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, y en tal caso definimos su *diferencial* como la aplicación lineal $df(c) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ cuyas funciones coordenadas son $df_1(c)$ y $df_2(c)$, es decir,

$$df(c)(v) = (df_1(c)(v), df_2(c)(v)) = \left(\frac{\partial f_1}{\partial x} \Big|_c v_1 + \frac{\partial f_1}{\partial y} \Big|_c v_2, \frac{\partial f_2}{\partial x} \Big|_c v_1 + \frac{\partial f_2}{\partial y} \Big|_c v_2 \right)$$

o, equivalentemente, recordando que toda aplicación lineal se puede expresar en términos de una matriz: $df(c)(v) = vJf(c)$, donde

$$Jf(c) = \begin{pmatrix} \frac{\partial f_1}{\partial x} \Big|_c & \frac{\partial f_2}{\partial x} \Big|_c \\ \frac{\partial f_1}{\partial y} \Big|_c & \frac{\partial f_2}{\partial y} \Big|_c \end{pmatrix}$$

es la *matriz jacobiana* de f en el punto c .

Por ejemplo, en el caso de las coordenadas polares tenemos las funciones coordenadas

$$x(\rho, \theta) = \rho \cos \theta, \quad y(\rho, \theta) = \rho \operatorname{sen} \theta,$$

que son diferenciables y sus diferenciales son

$$dx = \cos \theta d\rho - \rho \operatorname{sen} \theta d\theta, \quad dy = \operatorname{sen} \theta d\rho + \rho \cos \theta d\theta,$$

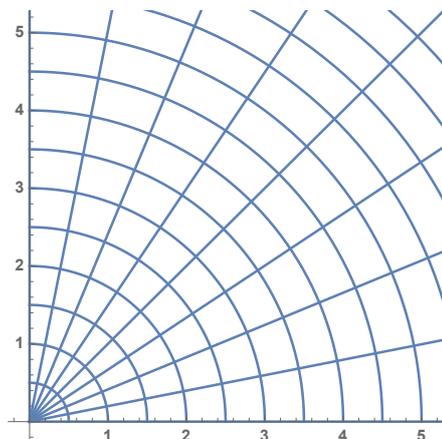
luego la función vectorial $P(\rho, \theta)$ también es diferenciable, y su diferencial es

$$dP = (\cos \theta d\rho - \rho \operatorname{sen} \theta d\theta, \operatorname{sen} \theta d\rho + \rho \cos \theta d\theta),$$

con matriz jacobiana:

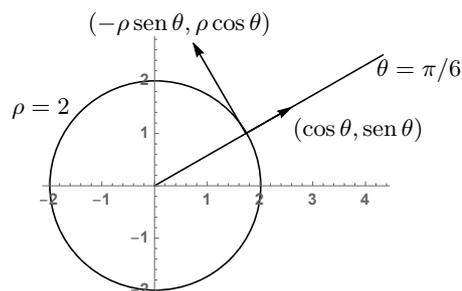
$$JP = \begin{pmatrix} \cos \theta & \operatorname{sen} \theta \\ -\rho \operatorname{sen} \theta & \rho \cos \theta \end{pmatrix}.$$

Cuando consideramos coordenadas cartesianas (x, y) , las curvas $x = x_0$ son rectas verticales, y las curvas $y = y_0$ son rectas horizontales, y todas ellas “cuadrículan” el plano, en el sentido de que todo desplazamiento puede descomponerse en un desplazamiento horizontal seguido de uno vertical, o viceversa, siguiendo “líneas de la cuadrícula cartesiana”. Cuando consideramos coordenadas polares la situación es parecida, solo que la “cuadrícula” es curva. La figura siguiente muestra varias líneas $\rho = \rho_0$, que son circunferencias de centro en el origen de coordenadas y varias líneas $\theta = \theta_0$, que son semirrectas que parten del origen de coordenadas:



Todo desplazamiento (entre puntos distintos del origen), puede realizarse mediante un movimiento radial seguido de uno transversal (circular) o viceversa.

En general, para una aplicación $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^2$, las curvas $f(x, y_0)$ o $f(x_0, y)$ (donde varían, respectivamente x e y , mientras y_0, x_0 permanecen constantes) definen una “cuadrícula curvilínea”, cuyos vectores tangentes son los vectores de derivadas parciales de f respecto de la variable que no queda fija, pero éstos son precisamente las filas de la matriz jacobiana.



Por ejemplo, en el caso de las coordenadas polares, es obvio que cada curva $\rho = \rho_0$ es ortogonal a cada curva $\theta = \theta_0$, lo cual se corresponde con el hecho de que

$$(-\rho \operatorname{sen} \theta, \rho \operatorname{cos} \theta) \cdot (\operatorname{cos} \theta, \operatorname{sen} \theta) = 0.$$

En general sabemos que $T_i(x) = f_i(c) + df_i(c)(x - c)$ es una buena aproximación de $f_i(x)$ para puntos x cercanos a c , por lo que

$$T(x) = f(x) + df(c)(x - c)$$

es una buena aproximación de $f(x)$ para puntos x cercanos a c .

En particular, si f es una aplicación lineal,

$$f(x, y) = (ax + by, cx + dy) = (x, y) \begin{pmatrix} a & c \\ b & d \end{pmatrix},$$

derivando obtenemos que su matriz jacobiana es precisamente la matriz de f , lo que significa que $df(c) = f$, para todo punto c . Esto expresa que “la mejor aproximación lineal de una aplicación lineal es ella misma”. Más en particular, es claro que la matriz jacobiana de la aplicación identidad $f(x, y) = (x, y)$ es la matriz identidad.

Observemos ahora que la regla de la cadena se generaliza trivialmente a funciones vectoriales:³

Teorema 5.15 (Regla de la cadena vectorial) Sean

$$y : U \subset \mathbb{R}^2 \longrightarrow V \subset \mathbb{R}^2, \quad z : V \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^2,$$

dos aplicaciones diferenciables (resp. de clase C^∞) en sus dominios. Entonces, la función compuesta $z \circ y : U \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ dada por $z \circ y(x) = z(y(x))$ es diferenciable en U y

$$dz(x)(v) = dz(y(x))(dy(x)(v)).$$

En efecto, basta aplicar la regla de la cadena a las funciones coordenadas z_1 y z_2 .

En otras palabras, la diferencial en un punto x de la función que resulta de aplicar $y(x)$ y luego $z(y)$ es la aplicación lineal que sobre un vector v se calcula aplicándole $dy(x)$ y luego $dz(y(x))$.

Puesto que una diferencial sobre un vector se puede calcular multiplicando el vector por la matriz jacobiana, tenemos que

$$dz(x)(v) = vJy(x) \cdot J(z(y(x))),$$

pero esto significa que la matriz jacobiana de la función compuesta $z \circ y$ es el producto

$$Jz \circ y(x) = Jy(x) \cdot Jz(y(x))$$

de las matrices jacobianas de las funciones $y(x)$ y $z(y)$.

En particular, si tenemos dos funciones diferenciables mutuamente inversas:

$$f : U \subset \mathbb{R}^2 \longrightarrow V \subset \mathbb{R}^2, \quad f^{-1} : V \subset \mathbb{R}^2 \longrightarrow U \subset \mathbb{R}^2,$$

su composición es la identidad en U (o en V) luego, si $f(x) = y$, entonces

$$Jf(x)Jf^{-1}(y) = I,$$

donde I es la matriz identidad, lo que significa que $Jf^{-1}(y)$ es la matriz inversa de $Jf(x)$.

Vamos a comprobar este hecho en el caso de las coordenadas polares. Para que P tenga inversa en un abierto necesitamos restringirla a un conjunto como $U =]0, +\infty[\times]-\pi, \pi[$, de modo que su imagen es el abierto V formado por todos los puntos del plano menos la semirrecta formada por los puntos $(x, 0)$, con $x \leq 0$.

³Es obvio que todo lo que estamos diciendo vale para funciones vectoriales arbitrarias $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$, pero no vamos a necesitar más que el caso de funciones en \mathbb{R}^2 .

Aun así, la inversa $P^{-1} : V \rightarrow U$ se expresa de forma un tanto farragosa:

$$P^{-1}(x, y) = (\rho(x, y), \theta(x, y)),$$

donde $\rho(x, y) = \sqrt{x^2 + y^2}$ y

$$\theta(x, y) = \begin{cases} \arctan(y/x) & \text{si } x > 0, \\ \pi/2 - \arctan(x/y) & \text{si } y > 0, \\ -\arctan(x/y) - \pi/2 & \text{si } y < 0. \end{cases}$$

Notemos que los puntos que están en varios casos tienen asignado el mismo valor de $\theta(x, y)$ en todos ellos. Damos definiciones alternativas porque así es evidente que $\theta(x, y)$ (luego también $P(x, y)$) es diferenciable en V , pues, dado cualquier punto $(x, y) \in V$ en todos los puntos de su alrededor P^{-1} está definida por una misma de las tres expresiones, y cualquiera de ellas es trivialmente diferenciable (de clase C^∞ , en realidad).

El lector puede comprobar que las tres expresiones dan —como no podía ser de otro modo— el mismo valor para las derivadas parciales, con lo que

$$dP^{-1} = \left(\frac{x}{\sqrt{x^2 + y^2}} dx + \frac{y}{\sqrt{x^2 + y^2}} dy, -\frac{y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy \right),$$

y la matriz jacobiana es

$$JP^{-1} = \begin{pmatrix} \frac{x}{\sqrt{x^2 + y^2}} & -\frac{y}{x^2 + y^2} \\ \frac{y}{\sqrt{x^2 + y^2}} & \frac{x}{x^2 + y^2} \end{pmatrix}.$$

A su vez,

$$JP^{-1}(x(\rho, \theta), y(\rho, \theta)) = \begin{pmatrix} \cos \theta & -(1/\rho) \sin \theta \\ \sin \theta & (1/\rho) \cos \theta \end{pmatrix},$$

y así es fácil comprobar que se trata de la matriz inversa de $JP(\rho, \theta)$.

El hecho de que dos funciones inversas tienen diferenciales o matrices jacobianas inversas es una parte del contenido del teorema de la función inversa para funciones de varias variables, pero hay una parte más que no es trivial:

Teorema 5.16 (Teorema de la función inversa) *Sea $f : U \rightarrow V$ una función biyectiva de clase C^1 entre dos abiertos en \mathbb{R}^2 tal que $Jf(x)$ tenga determinante no nulo en cada punto. Entonces $f^{-1} : V \rightarrow U$ también es de clase C^1 , y es de clase C^∞ si f lo es. Además, si $f(x) = y$, se cumple que $Jf^{-1}(y)$ es la matriz inversa de $Jf(x)$.*

Lo único que no es evidente teniendo en cuenta lo que hemos visto hasta el momento es que la función f^{-1} es diferenciable. Admitiendo que lo es, sabemos que $Jf^{-1}(y)$ es la matriz inversa de $Jf(x)$ o, explícitamente, teniendo en cuenta la fórmula para la matriz inversa dada en [ITA] 11.7], es la matriz

$$\left(\begin{array}{cc} \frac{\partial f_1}{\partial x} \Big|_{f^{-1}(y)} & \frac{\partial f_2}{\partial y} \Big|_{f^{-1}(y)} \\ -\frac{\partial f_2}{\partial x} \Big|_{f^{-1}(y)} & \frac{\partial f_1}{\partial y} \Big|_{f^{-1}(y)} \end{array} \right)^{-1} \cdot \left(\begin{array}{cc} \frac{\partial f_2}{\partial y} \Big|_{f^{-1}(y)} & -\frac{\partial f_2}{\partial x} \Big|_{f^{-1}(y)} \\ -\frac{\partial f_1}{\partial y} \Big|_{f^{-1}(y)} & \frac{\partial f_1}{\partial x} \Big|_{f^{-1}(y)} \end{array} \right),$$

donde el primer término es el inverso del determinante de Jf , que por hipótesis no se anula.

Como f es de clase C^1 , sus derivadas parciales son continuas y, como f^{-1} es diferenciable, es continua. Por lo tanto, todas las derivadas que aparecen en la fórmula anterior son continuas, luego las derivadas parciales de f^{-1} son continuas, luego f^{-1} es de clase C^1 . Si f es de clase C^∞ , como sus derivadas son de clase C^1 y f^{-1} también, todas las derivadas de la fórmula anterior son de clase C^1 , luego f es de clase C^2 , y continuando este razonamiento concluimos que f^{-1} es de clase C^∞ .

La prueba de que f^{-1} es diferenciable la posponemos a la sección siguiente (página 321). Observemos que la hipótesis de que el determinante de la matriz jacobiana no se anule es necesaria, pues equivale a que la matriz tenga inversa, y hemos visto que si f^{-1} es diferenciable, entonces Jf tiene inversa.

En el caso de las coordenadas polares hemos podido comprobar directamente que la aplicación inversa es diferenciable, pero vamos a ver un ejemplo similar en el que la diferenciabilidad de la inversa no es trivial:

Coordenadas cicloidales Recordemos que la cicloide generada por una circunferencia de radio R es la curva

$$r(t) = (Rt - R \operatorname{sen} t, R(1 - \operatorname{cos} t)),$$

que parte de $(0, 0)$ y recorre un arco completo cuando $0 \leq t \leq 2\pi$.

Vamos a probar que podemos llegar a todo punto del cuadrante positivo siguiendo una única cicloide:

Teorema 5.17 *Si $x > 0$ e $y \geq 0$, existen unos únicos $R > 0$ y $0 \leq t \leq 2\pi$ tales que*

$$(x, y) = (Rt - R \operatorname{sen} t, R(1 - \operatorname{cos} t)).$$

DEMOSTRACIÓN: Observemos que si $0 < t < 2\pi$, se cumplen las desigualdades siguientes:

- a) $t - \operatorname{sen} t > 0$,
- b) $\operatorname{sen}(t/2) - (t/2) \operatorname{cos}(t/2) > 0$,
- c) $2(1 - \operatorname{cos} t) - t \operatorname{sen} t > 0$.

En efecto, si $f(t) = t - \operatorname{sen} t$, por el teorema del valor medio, si $0 < t < 2\pi$, tenemos que existe $0 < t_0 < t$ tal que

$$f(t) = f(t) - f(0) = f'(t_0)t > 0.$$

El mismo razonamiento prueba la segunda desigualdad, pues la función también tiene derivada positiva en $]0, 2\pi[$. Por último,

$$2(1 - \cos t) - t \operatorname{sen} t = 4 \operatorname{sen}^2 \frac{t}{2} - 2t \operatorname{sen} \frac{t}{2} \cos \frac{t}{2} = 4 \operatorname{sen} \frac{t}{2} \left(\operatorname{sen} \frac{t}{2} - \frac{t}{2} \cos \frac{t}{2} \right) > 0,$$

donde usamos la segunda desigualdad.

A continuación consideramos la función

$$h(t) = \frac{1 - \cos t}{t - \operatorname{sen} t},$$

que está definida en $]0, 2\pi]$ y su derivada es

$$h'(t) = \frac{2(\cos t - 1) + t \operatorname{sen} t}{(t - \operatorname{sen} t)^2} < 0$$

en $]0, 2\pi[$ por la tercera desigualdad precedente. Por lo tanto h es estrictamente decreciente y $h(2\pi) = 0$, mientras que, por la regla de L'Hôpital,

$$\begin{aligned} \lim_{t \rightarrow 0^+} h(t) &= \lim_{t \rightarrow 0^+} \frac{\operatorname{sen} t}{-2(t - \operatorname{sen} t) \cos t} = \lim_{t \rightarrow 0^+} \frac{\tan t}{-2(t - \operatorname{sen} t)} \\ &= \lim_{t \rightarrow 0^+} \frac{1 + \tan^2 t}{-2(1 - \cos t)} = +\infty. \end{aligned}$$

Por el teorema de los valores intermedios, existe un único $0 < t \leq 2\pi$ tal que

$$\frac{y}{x} = h(t) = \frac{1 - \cos t}{t - \operatorname{sen} t}$$

y, llamando

$$R = \frac{x}{t - \operatorname{sen} t},$$

se cumple que $(x, y) = (Rt - R \operatorname{sen} t, R(1 - \cos t))$. Además R y t son únicos, pues si R^*, t^* cumplen lo mismo, tenemos que $y/x = h(t^*)$, luego tiene que ser $t^* = t$, y entonces $R(t - \operatorname{sen} t) = x = R^*(t - \operatorname{sen} t) \neq 0$, luego $R^* = R$. ■

Ahora llamamos $U = \{(\rho, \tau) \in \mathbb{R}^2 \mid 0 < \tau < 2\pi\rho\}$ y definimos $f : U \rightarrow \mathbb{R}^2$ mediante

$$f(\rho, \tau) = (\rho\tau - \rho^2 \operatorname{sen}(\tau/\rho), \rho^2(1 - \cos(\tau/\rho))),$$

que resulta de hacer

$$R = \rho^2, \quad t = \tau/\rho$$

en la fórmula de la cicloide. Si $V =]0, +\infty[\times]0, +\infty[$, el teorema anterior prueba que $f : U \rightarrow V$ es biyectiva.

La aplicación f es obviamente de clase C^∞ en U , y su matriz jacobiana es

$$Jf = \begin{pmatrix} \tau + \tau \cos \frac{\tau}{\rho} - 2\rho \operatorname{sen} \frac{\tau}{\rho} & 2\rho(1 - \cos \frac{\tau}{\rho}) - \tau \operatorname{sen} \frac{\tau}{\rho} \\ \rho - \rho \cos \frac{\tau}{\rho} & \rho \operatorname{sen} \frac{\tau}{\rho} \end{pmatrix}.$$

Un cálculo rutinario muestra que el determinante es

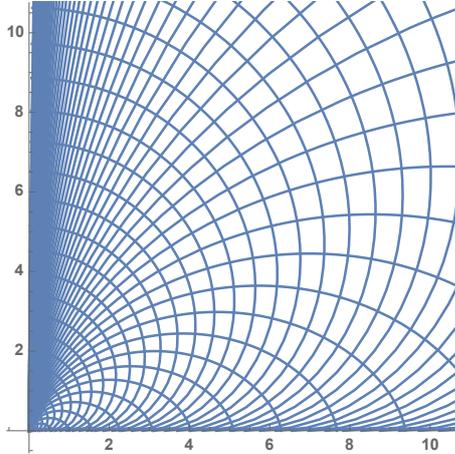
$$-2\rho^2 \left(2(1 - \cos \frac{\tau}{\rho}) - \frac{\tau}{\rho} \operatorname{sen} \frac{\tau}{\rho} \right),$$

y esta expresión es < 0 por la desigualdad c) probada en el teorema anterior. Por lo tanto, el teorema de la función inversa nos asegura que f^{-1} también es de clase C^∞ , cosa que no es fácil de comprobar directamente, pues no es posible despejar las funciones $\rho(x, y)$ y $\tau(x, y)$.

Sin embargo, aunque no sepamos calcularlas explícitamente, podemos considerar a $\rho(x, y)$, $\theta(x, y)$ como unas coordenadas alternativas para cualquier punto del cuadrante positivo V , de modo que el punto de coordenadas cicloidales (ρ, τ) , es decir, el punto con coordenadas cartesianas

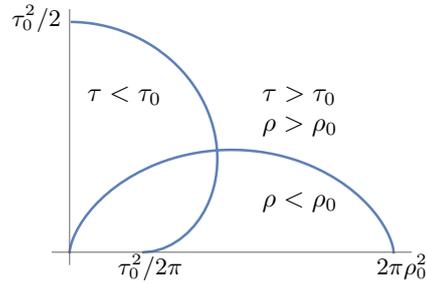
$$x = \rho\tau - \rho^2 \operatorname{sen}(\tau/\rho), \quad y = \rho^2(1 - \cos(\tau/\rho)),$$

es el punto situado sobre la cicloide generada por una circunferencia de radio ρ^2 correspondiente a un giro de τ/ρ radianes. La figura siguiente muestra la “cuadrícula curvilínea” correspondiente:



Las curvas $\rho = \rho_0$ son cicloides, y las curvas $\tau = \tau_0$ cortan ortogonalmente a cada cicloide, como muestra la figura y como podemos comprobar multiplicando las filas de la matriz jacobiana y comprobando que el resultado es 0.

Más precisamente, una curva $\rho = \rho_0$ empieza en $(0, 0)$, cuando $\tau = 0$, y termina en $(2\pi\rho_0^2, 0)$, cuando $\tau = 2\pi\rho_0$, mientras que una curva $\tau = \tau_0$ empieza en $(\tau_0^2/2\pi, 0)$, cuando $\rho = \tau_0/2\pi$, y termina en $(\tau_0^2/2, 0)$ cuando ρ tiende a $+\infty$.



En efecto, basta tener en cuenta que, por los desarrollos en serie de potencias,

$$\sin x = x + x^3 f(x), \quad \cos x = 1 + x^2 g(x),$$

donde f y g son funciones continuas en \mathbb{R} con $f(0) = -1/6$, $g(0) = -1/2$. Por lo tanto,

$$\lim_{\rho \rightarrow +\infty} \rho\tau_0 - \rho^2 \sin \frac{\tau_0}{\rho} = \lim_{\rho \rightarrow +\infty} \rho\tau_0 - \rho^2 \frac{\tau_0}{\rho} + \rho^2 \frac{\tau_0^3}{\rho^3} f(\tau_0/\rho) = \lim_{\rho \rightarrow +\infty} \frac{\tau_0^3}{\rho} f(\tau_0/\rho) = 0,$$

$$\lim_{\rho \rightarrow +\infty} \rho^2 (1 - \cos \frac{\tau_0}{\rho}) = \lim_{\rho \rightarrow +\infty} -\rho^2 \frac{\tau_0^2}{\rho^2} g(\tau_0/\rho) = \frac{\tau_0^2}{2}.$$

Más aún, observamos que los puntos con coordenada $\rho < \rho_0$ son los que están debajo de la cicloide $\rho = \rho_0$, y los puntos con $\tau < \tau_0$ son los que están a la izquierda de la curva $\tau = \tau_0$. De aquí se sigue una última propiedad que vamos a necesitar:

Teorema 5.18 Si $\{(x_n, y_n)\}_{n=0}^{\infty}$ es una sucesión de puntos con $x_n > 0$, $y_n > 0$ tal que

$$\lim_n (x_n, y_n) = (0, y),$$

con $y > 0$, entonces la sucesión $\{(\rho_n, \tau_n)\}_{n=0}^{\infty}$ de sus coordenadas cicloidales cumple

$$\lim_n \rho_n = +\infty, \quad \lim_n \tau_n = \sqrt{2y}.$$

Similarmente, si

$$\lim_n (x_n, y_n) = (x, 0),$$

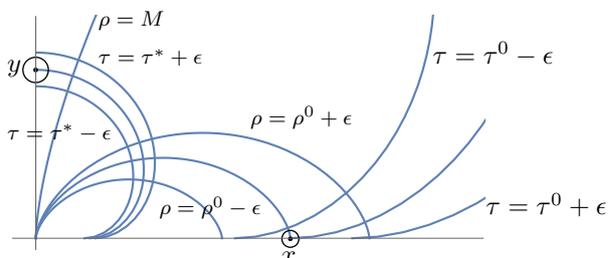
con $x > 0$, entonces

$$\lim_n \rho_n = \sqrt{\frac{x}{2\pi}}, \quad \lim_n \tau_n = \sqrt{2\pi x}.$$

DEMOSTRACIÓN: Dado cualquier $M > 0$, podemos tomar una bola centrada en $(0, y)$ que esté fuera de la cicloide $\rho = M$ (véase la figura), luego, si la sucesión dada converge a $(0, y)$, para todo n suficientemente grande, sus puntos estarán en dicha bola, luego cumplirán $\rho_n > M$. Esto prueba que ρ_n tiende a $+\infty$.

Similarmente, sea $\tau^* = \sqrt{2y}$, de modo que la curva $\tau = \tau^*$ termina en el punto $(0, y)$. Entonces, dado $0 < \epsilon < \tau^*$, la curva $\tau = \tau^* + \epsilon$ termina

en $(\tau^* + \epsilon)^2/2 > y$, mientras que la curva $\tau = \tau^* - \epsilon < y$, luego podemos tomar una bola de centro $(0, y)$ comprendida entre ambas, y así, para todo n suficientemente grande, se tiene que cumplir que $\tau^* - \epsilon < \tau_n < \tau^* + \epsilon$, lo que prueba que τ_n tiende a τ^* .



La segunda parte del teorema se demuestra análogamente, considerando ahora $\rho^0 = \sqrt{x/2\pi}$, $\tau^0 = \sqrt{2\pi x}$ y una bola centrada en $(x, 0)$ comprendida entre las curvas $\rho = \rho^0 - \epsilon$, $\rho = \rho^0 + \epsilon$, $\tau = \tau^0 - \epsilon$, $\tau = \tau^0 + \epsilon$. ■

El vector gradiente Ya hemos discutido todos los resultados de enunciado natural y prueba técnica que vamos a necesitar para manejar cómodamente las funciones diferenciables. En el estudio de las coordenadas cicloidales ya hemos visto un ejemplo de lo que supone trabajar con funciones diferenciables sin entrar en tecnicismos. Terminamos esta sección con un segundo ejemplo:

Definición 5.19 Si $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es una función diferenciable su *vector gradiente* en un punto $c \in A$ es el vector de derivadas parciales⁴

$$\nabla f(c) = \left(\left. \frac{\partial f}{\partial x_1} \right|_c, \dots, \left. \frac{\partial f}{\partial x_n} \right|_c \right).$$

De hecho, podemos considerar que $\nabla f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$.

Aunque el razonamiento siguiente es válido en general, nos restringimos a $n = 3$ para no salirnos del caso en el que conocemos su interpretación geométrica (y todo vale trivialmente si $n = 2$):

Si $f : A \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ es una función diferenciable, su derivada direccional en un punto $c \in A$ en una dirección v que podemos suponer de norma 1 viene dada por

$$f'(c; v) = df(c)(v) = \left. \frac{\partial f}{\partial x} \right|_c v_1 + \left. \frac{\partial f}{\partial y} \right|_c v_2 + \left. \frac{\partial f}{\partial z} \right|_c v_3 = \nabla f(c) \cdot v = \|\nabla f(c)\| \cos \alpha,$$

donde α es el ángulo que forman v y $\nabla f(c)$. Suponiendo que el gradiente sea no nulo, la máxima derivada direccional se dará cuando $\alpha = 0$, es decir, cuando v tenga la dirección del gradiente, y su valor será $\|\nabla f(c)\|$. Así pues:

⁴El signo ∇ fue usado por primera vez por Hamilton, aunque no le dio nombre, y más adelante, Peter Guthrie Tait y James Clerk Maxwell, en una conversación en la que no sabían cómo llamarlo, acabaron llamándolo “nabla” en broma, que es una palabra griega que nombraba a un cierto tipo de arpa, pero el nombre “cuajó” y así es como se le llama ahora.

El gradiente de una función diferenciable, si es no nulo, indica la dirección de máximo crecimiento de la función, es decir, la dirección en la que la derivada direccional es máxima, y su valor es precisamente la norma del gradiente.

Similarmente, la derivada direccional más negativa de f se obtiene cuando $\alpha = 180^\circ$, luego es la dirección opuesta al gradiente, que es, por tanto, la *dirección de máximo decrecimiento* de f .

Por otra parte, las direcciones ortogonales al gradiente son *direcciones de crecimiento nulo*, es decir, direcciones en las que la derivada direccional es nula. Naturalmente, si el gradiente es nulo todas las derivadas direccionales son nulas, lo cual puede deberse a que la función alcanza en c un valor máximo o mínimo, si bien no es necesario que así sea, como hemos visto en el ejemplo de la pág. 289.

Si $r(t)$ es una curva que pasa por $r(t_0) = c$ y sobre la que f toma un valor constante, entonces, por la regla de la cadena,

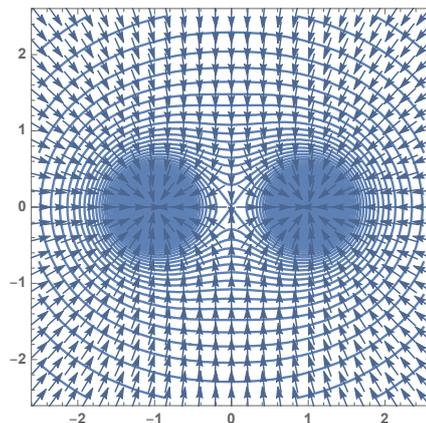
$$0 = \frac{df(r(t))}{dt} = \frac{df}{dx} \frac{dx}{dt} + \frac{df}{dy} \frac{dy}{dt} + \frac{df}{dz} \frac{dz}{dt} = \nabla f \cdot \frac{dr}{dt},$$

luego el gradiente es ortogonal a la trayectoria.

Ejemplo La figura muestra varias curvas donde la función

$$f(x, y) = \frac{1}{\sqrt{(x+1)^2 + y^2}} + \frac{1}{\sqrt{(x-1)^2 + y^2}}$$

toma un valor constante, junto con varias flechas que apuntan en la dirección del gradiente en cada punto (aunque su longitud es constante para que el dibujo no quede confuso):



La función f tiende a $+\infty$ en $(\pm 1, 0)$ y, en efecto, vemos cómo los gradientes marcan caminos que señalan hacia dichos puntos y son ortogonales a las curvas. ■

5.3 Teoría básica

Dedicamos esta sección a demostrar los resultados que, por su carácter técnico, hemos dejado pendientes en la sección anterior.

DEMOSTRACIÓN DE 5.6: Si $f : A \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es diferenciable en $c \in A$, entonces

$$\lim_{v \rightarrow 0} \frac{f(c+v) - f(c) - df(c)(v)}{\|v\|} = 0.$$

Esto quiere decir que, dado $\epsilon > 0$, existe un $\eta > 0$ tal que si $0 < \|v\| < \eta$, entonces

$$|f(c+v) - f(c) - df(c)(v)| < \|v\| \sqrt{\frac{\epsilon}{2}}.$$

Si llamamos

$$M = \max\left\{\left|\frac{\partial f}{\partial x_1}\right|_c, \dots, \left|\frac{\partial f}{\partial x_n}\right|_c\right\},$$

entonces

$$\begin{aligned} |f(c+v) - f(c)| &\leq |f(c+v) - f(c) - df(c)(v)| + |df(c)(v)| \\ &< \|v\| \sqrt{\frac{\epsilon}{2}} + \sum_{i=1}^n \left|\frac{\partial f}{\partial x_i}\right|_c |v_i| \leq \|v\| \sqrt{\frac{\epsilon}{2}} + nM\|v\|. \end{aligned}$$

Por lo tanto, si $\delta = \min\{\sqrt{\epsilon/2}, \epsilon/2nM\}$ y $\|v\| < \delta$, tenemos que

$$|f(c+v) - f(c)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Equivalentemente, si $\|x-c\| < \delta$, se cumple que $|f(x) - f(c)| < \epsilon$, y esto significa que f es continua en c . ■

DEMOSTRACIÓN DE 5.7: Sea $c \in A$ y vamos a probar que f es diferenciable en c . Para ello basta probar que

$$\lim_{v \rightarrow 0} \frac{f(c+v) - f(c) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}\Big|_c v_i}{\|v\|} = 0,$$

lo que a su vez equivale a que, dado $\epsilon > 0$, exista un $\delta > 0$ de modo que si $0 < \|v\| < \delta$ entonces

$$\left|f(c+v) - f(c) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}\Big|_c v_i\right| < \epsilon\|v\|.$$

Por la continuidad de las derivadas parciales tenemos que existe un $\delta > 0$ tal que si $\|y-c\| < \delta$ entonces $y \in A$ (porque A es abierto) y

$$\left|\frac{\partial f}{\partial x_i}\Big|_y - \frac{\partial f}{\partial x_i}\Big|_c\right| < \frac{\epsilon}{n}$$

para $i = 1, \dots, n$. Fijemos un v tal que $\|v\| < \delta$.

Queremos probar que la variación que experimenta f cuando, a partir del punto c , modificamos cada una de sus variables en la cantidad v_i , la variación se aproxima adecuadamente con la suma del producto de cada derivada parcial por la variación correspondiente. Vamos a ver que podemos reducir el problema a estudiar lo que sucede cuando sólo se modifica una variable. Para ello definimos

$$F_i = f(c_1 + v_1, \dots, c_i + v_i, c_{i+1}, \dots, c_n),$$

es decir, el valor de f cuando modificamos sólo las primeras i -variables.

En particular, vemos que $f(c + v) = F_n$ y $f(c) = F_0$, luego

$$\begin{aligned} \left| f(c + v) - f(c) - \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_c v_i \right| &= \left| \sum_{i=1}^n (F_i - F_{i-1} - \frac{\partial f}{\partial x_i} \Big|_c v_i) \right| \\ &\leq \sum_{i=1}^n \left| F_i - F_{i-1} - \frac{\partial f}{\partial x_i} \Big|_c v_i \right|, \end{aligned}$$

pues en la segunda suma tenemos $F_1 - F_0 + F_2 - F_1 + F_3 - F_2 + \dots$ y se cancelan todos los términos menos $-F_0$ al principio y F_n al final. Así pues, (teniendo en cuenta que $|v_i| \leq \|v\|$) basta probar que

$$\left| F_i - F_{i-1} - \frac{\partial f}{\partial x_i} \Big|_c v_i \right| < \frac{\epsilon}{n} |v_i|,$$

para $i = 1, \dots, n$.

Así, lo que tenemos que demostrar es que la variación de f que resulta de modificar únicamente la variable i -ésima en una cantidad v_i se aproxima adecuadamente por el producto de la parcial i -ésima por el incremento v_i . El único detalle es que la derivada está calculada en c y ahora el punto de partida del incremento no es c , sino c con las primeras $i - 1$ variables ya incrementadas. Vamos a aplicar el teorema del valor medio a la función de una variable dada por

$$g_i(t) = f(c_1 + v_1, \dots, c_{i-1} + v_{i-1}, c_i + tv_i, c_{i+1}, \dots, c_n).$$

Observemos que esta función va calculando la variación de f cuando partimos del punto c con las primeras $i - 1$ variables incrementadas y vamos incrementando gradualmente la i -ésima. Par $t = 0$ estamos en c_i y para $t = 1$ estamos en $c_i + v_i$.

Esta función está definida en un intervalo abierto que contiene a $[0, 1]$, y el hecho de que f tenga derivada parcial i -ésima en A implica que la función

$$h(x) = f(c_1 + v_1, \dots, c_{i-1} + v_{i-1}, x, c_{i+1}, \dots, c_n)$$

es derivable en los puntos x tales que $(c_1 + v_1, \dots, c_{i-1} + v_{i-1}, x, c_{i+1}, \dots, c_n)$ está en A , luego $g_i(t) = h(c_i + tv_i)$ es derivable, luego también continua, en su dominio. En particular es continua en $[0, 1]$ y derivable en $]0, 1[$. Además, por la regla de la cadena, $g'_i(t) = h'(c_i + tv_i)v_i$, es decir:

$$g'_i(t) = \frac{\partial f}{\partial x_i} \Big|_{(a_1+v_1, \dots, a_{i-1}+v_{i-1}, a_i+tv_i, a_{i+1}, \dots, a_n)} v_i.$$

El teorema del valor medio nos da que existe $0 < t_0 < 1$ tal que

$$F_i - F_{i-1} = g_i(1) - g_i(0) = g'_i(t_0)(1 - 0) = \left. \frac{\partial f}{\partial x_i} \right|_y v_i,$$

donde $y = (c_1 + v_1, \dots, c_{i-1} + v_{i-1}, c_i + t_0 v_i, c_{i+1}, \dots, c_n)$ cumple

$$\|y - c\| = \|(v_1, \dots, v_{i-1}, t_0 v_i, 0, \dots, 0)\| \leq \|v\| < \delta,$$

luego, por la continuidad de la derivada parcial:

$$\left| F_i - F_{i-1} - \left. \frac{\partial f}{\partial x_i} \right|_c v_i \right| = \left| \left. \frac{\partial f}{\partial x_i} \right|_y - \left. \frac{\partial f}{\partial x_i} \right|_c \right| |v_i| < \frac{\epsilon}{n} |v_i|,$$

como había que probar. ■

Observaciones sobre la continuidad en \mathbb{R}^n Tal y como hemos señalado en la prueba del teorema 5.10, los apartados c) y d) requieren el hecho de que las funciones continuas tienen las propiedades enunciadas, cosa que en [ITAn] demostramos para funciones de una y dos variables. Concretamente, necesitamos la generalización de [ITAn 3.3] y el hecho de que la composición de funciones continuas es continua.

Aunque se puede seguir un camino más directo, si seguimos el enfoque de [ITAn], necesitamos considerar la generalización obvia de [ITAn 1.1] (la definición de convergencia de sucesiones). Con ella podemos probar la caracterización de la continuidad por sucesiones [ITAn 3.2] a partir de la cual es inmediata la generalización de [ITAn 3.3].

La propiedad d) del teorema 5.10 también es inmediata: si una sucesión $\{x_n\}_{n=0}^\infty$ converge a un punto $x \in A$, entonces $\{f(x_n)\}_{n=0}^\infty$ converge a $f(x)$ (por la generalización de [ITAn 3.2]), luego $\{g(f(x_n))\}_{n=0}^\infty$ converge a $g(f(x))$ (por [ITAn 3.2]), lo que prueba la continuidad de $g(f(x))$ (por la generalización de [ITAn 3.2]). No necesitamos nada más. ■

DEMOSTRACIÓN DEL TEOREMA DE SCHWARZ: No perdemos generalidad si suponemos que $n = 2$, pues podemos trabajar con la función

$$F(x_i, x_j) = f(a_1, \dots, x_i, \dots, x_j, \dots, a_n).$$

Fijemos un punto $c = (a, b)$. Como A es abierto, existe un $\eta > 0$ tal que si $\|(u, v)\| < \eta$ entonces $(a + u, b + v) \in A$. En particular, si $\delta = \eta/\sqrt{2}$, tenemos que si $h \in]-\delta, \delta[$ se cumple que (h, h) , $(h, 0)$ y $(0, h)$ tienen todos norma menor que η , luego está definida la función

$$\Delta f(h) = f(a + h, b + h) - f(a + h, b) - f(a, b + h) + f(a, b).$$

Vamos a probar que

$$\left. \frac{\partial^2 f}{\partial x \partial y} \right|_{(a,b)} = \lim_{h \rightarrow 0^+} \frac{\Delta f(h)}{h^2},$$

pero, como la definición de $\Delta f(h)$ es simétrica, lo mismo valdrá para la segunda derivada en orden inverso, con lo que ambas serán iguales y el teorema estará probado.

Dado $\epsilon > 0$, la continuidad de la derivada de orden 2 en (a, b) implica que, restringiendo η (y por consiguiente δ) si es preciso, podemos suponer que si $\|(k, k')\| < \eta$ (y en particular si $0 < k, k' < \delta$) se cumple

$$\left| \frac{\partial^2 f}{\partial x \partial y} \Big|_{(a+k, b+k')} - \frac{\partial^2 f}{\partial x \partial y} \Big|_{(a, b)} \right| < \epsilon.$$

Ahora, si $0 < h < \delta$, consideramos la función

$$G(t) = f(t, b+h) - f(t, b),$$

definida en el intervalo $[a, a+h]$. Por el teorema del valor medio existe un número $0 < k < h$ tal que

$$\begin{aligned} \Delta f(h) &= G(a+h) - G(a) = h G'(a+k) \\ &= h \left(\frac{\partial f}{\partial x} \Big|_{(a+k, b+h)} - \frac{\partial f}{\partial x} \Big|_{(a+k, b)} \right). \end{aligned}$$

Ahora aplicamos el teorema del valor medio a la función

$$H(t) = \frac{\partial f}{\partial x} \Big|_{(a+k, t)}$$

en el intervalo $[a, a+h]$, que nos da un número $0 < k' < h$ tal que

$$\frac{\partial f}{\partial x} \Big|_{(a+k, b+h)} - \frac{\partial f}{\partial x} \Big|_{(a+k, b)} = \frac{\partial^2 f}{\partial x \partial y} \Big|_{(a+k, b+k')} h.$$

En total tenemos que

$$\Delta f(h) = \frac{\partial^2 f}{\partial x \partial y} \Big|_{(a+k, b+k')} h^2$$

luego

$$\left| \frac{\Delta f(h)}{h^2} - \frac{\partial^2 f}{\partial x \partial y} \Big|_{(a, b)} \right| = \left| \frac{\partial^2 f}{\partial x \partial y} \Big|_{(a+k, b+k')} - \frac{\partial^2 f}{\partial x \partial y} \Big|_{(a, b)} \right| < \epsilon,$$

siempre que $0 < h < \delta$. ■

DEMOSTRACIÓN DE LA REGLA DE LA CADENA: Supongamos en primer lugar que las funciones dadas son meramente diferenciables. Fijemos un punto $x \in B$ y un vector $v \in \mathbb{R}^m$ tal que $x+v \in B$. Entonces

$$z(y(x+v)) - z(y(x)) = z(y+u) - z(y),$$

donde $y = y(x)$, $u = y(x+v) - y(x)$.

La diferenciabilidad en x de las funciones y_i y la diferenciabilidad en y de la función z equivalen a que las funciones

$$E_i(v) = \frac{y_i(x+v) - y_i(x) - dy_i(x)(v)}{\|v\|}, \quad F(u) = \frac{z(y+u) - z(y) - dz(y)(u)}{\|u\|}$$

tienden a 0 cuando v y u tienden a 0, respectivamente. Despejando:

$$z(y(x+v)) - z(y(x)) = dz(y)(u) + \|u\|F(u),$$

$$u_i = y_i(x+v) - y_i(x) = dy_i(x)(v) + \|v\|E_i(v),$$

luego al sustituir queda:

$$\begin{aligned} z(y(x+v)) - z(y(x)) &= dz(y)(dy_1(x)(v), \dots, dy_n(x)(v)) \\ &+ \|v\|dz(y)(E_1(v), \dots, E_n(v)) + \|y(x+v) - y(x)\|F(y(x+v) - y(x)) \\ &= \phi(v) + \|v\|E(v), \end{aligned}$$

donde

$$\begin{aligned} \phi(v) &= dz(y(x))(dy_1(x)(v), \dots, dy_n(x)(v)) = \sum_{i=1}^n \frac{\partial z}{\partial y_i} \Big|_{y(x)} \sum_{j=1}^m \frac{\partial y_i}{\partial x_j} \Big|_x v_j \\ &= \sum_{j=1}^m \left(\sum_{i=1}^n \frac{\partial z}{\partial y_i} \Big|_{y(x)} \frac{\partial y_i}{\partial x_j} \Big|_x \right) v_j \end{aligned}$$

es una aplicación lineal y

$$E(v) = dz(y)(E_1(v), \dots, E_n(v)) + \frac{\|y(x+v) - y(x)\|}{\|v\|} F(y(x+v) - y(x)).$$

Basta probar que $\lim_{v \rightarrow 0} E(v) = 0$, pues esto significa que

$$\lim_{v \rightarrow 0} \frac{z(y(x+v)) - z(y(x)) - \phi(v)}{\|v\|} = 0,$$

con lo que $z(y(x))$ será diferenciable en x con diferencial $\phi(v)$, como queremos probar. Veamos que los dos sumandos que definen E tienden a 0. Para el primero, llamamos

$$M = \max \left\{ \left| \frac{\partial z}{\partial y_1} \Big|_y \right|, \dots, \left| \frac{\partial z}{\partial y_n} \Big|_y \right| \right\}$$

y observamos que, dado $\epsilon > 0$,

$$|dz(y)(E_1(v), \dots, E_n(v))| = \left| \sum_{i=1}^n \frac{\partial z}{\partial y_i} \Big|_y E_i(v) \right| \leq \sum_{i=1}^n M |E_i(v)|$$

luego tomando v de norma suficientemente pequeña como para que se cumpla $|E_i(v)| < \epsilon/(nM)$, tenemos que $|dz(y)(E_1(v), \dots, E_n(v))| < \epsilon$.

Para el segundo llamamos

$$M'_i = \max\left\{\left|\frac{\partial y_i}{\partial x_1}\right|_x, \dots, \left|\frac{\partial y_i}{\partial x_m}\right|_x\right\},$$

de modo que, si tomamos v de norma suficientemente pequeña como para que $|E_i(v)| \leq 1$, tenemos:

$$\begin{aligned} |y_i(x+v) - y_i(x)| &\leq |dy_i(x)(v)| + \|v\| |E_i(v)| \leq \left|\sum_{j=1}^m \frac{\partial y_i}{\partial x_j}\right|_x v_j + \|v\| \\ &\leq \sum_{j=1}^m M'_i \|v\| + \|v\| \leq (mM' + 1)\|v\|, \end{aligned}$$

y de aquí que

$$\|y(x+v) - y(x)\| = \sqrt{\sum_{i=1}^n |y_i(x+v) - y_i(x)|^2} \leq \sqrt{n(mM' + 1)\|v\|^2} = K\|v\|,$$

donde $K = \sqrt{n(mM' + 1)}$. Por lo tanto

$$\left|\frac{\|y(x+v) - y(x)\|}{\|v\|} F(y(x+v) - y(x))\right| \leq K|F(y(x+v) - y(x))|.$$

Como $\lim_{u \rightarrow 0} F(u) = 0$, dado $\epsilon > 0$, existe un $\delta > 0$ tal que si $\|u\| < \delta$ entonces $|F(u)| < \epsilon/K$, luego si $\|v\| < \delta/K$ se cumple que $\|y(x+v) - y(x)\| < \delta$, luego $K|F(y(x+v) - y(x))| < \epsilon$ y así queda probado que el segundo sumando también tiende a 0.

Con esto queda probada la regla de la cadena para funciones diferenciables. Si además las funciones son de clase C^∞ , fórmula (5.5) que se deduce de la parte ya probada implica que las derivadas de la función $z(x)$ también son de clase C^∞ , luego z es de clase C^∞ . ■

El teorema de la función inversa Necesitamos algunas consideraciones previas. Recordemos que hemos definido que una función f entre abiertos de \mathbb{R}^2 es diferenciable si lo son sus funciones coordenadas f_1 y f_2 . Explícitamente, la diferenciabilidad en un punto c equivale a que existan constantes p, q, r, s tales que

$$\begin{aligned} \lim_{v \rightarrow (0,0)} \frac{f_1(c+v) - f_1(c) - pv_1 - qv_2}{\|v\|} &= 0, \\ \lim_{v \rightarrow (0,0)} \frac{f_2(c+v) - f_2(c) - rv_1 - sv_2}{\|v\|} &= 0, \end{aligned}$$

pero observemos que

$$(pv_1 + qv_2, rv_1 + sv_2) = v \cdot M,$$

donde

$$M = \begin{pmatrix} p & r \\ q & s \end{pmatrix}$$

por lo que la condición puede expresarse vectorialmente así:

$$\lim_{v \rightarrow (0,0)} \frac{f(c+v) - f(c) - v \cdot M}{\|v\|} = 0.$$

La función f es diferenciable en c si y sólo si existe una matriz M con la cual se cumple la igualdad anterior.

En segundo lugar necesitamos el hecho siguiente:

Teorema 5.20 Sea $f : A \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$ una función diferenciable tal que, existen $\delta, K > 0$ de modo que si $\|x - x_0\| < \delta$, entonces $x \in A$ y

$$\left| \frac{\partial f_i}{\partial x_j} \right| \leq K, \quad i, j = 1, 2.$$

Entonces, si $\|a - x_0\|, \|b - x_0\| < \delta$, se cumple que $\|f(b) - f(a)\| \leq 3K\|b - a\|$.

DEMOSTRACIÓN: Fijados $a, b \in \mathbb{R}^2$ tales que $\|a - x_0\|, \|b - x_0\| < \delta$, si $0 \leq t \leq 1$, se cumple que

$$\begin{aligned} \|(1-t)a + tb - x_0\| &= \|(1-t)(a - x_0) + t(b - x_0)\| \leq \\ &(1-t)\|a - x_0\| + t\|b - x_0\| < (1-t)\delta + t\delta = \delta. \end{aligned}$$

Por lo tanto podemos definir

$$h_i(t) = f_i((1-t)a + tb).$$

Por la regla de la cadena

$$h'_i(t) = \frac{\partial g_i}{\partial x_1} \Big|_{(1-t)a+tb} (b_1 - a_1) + \frac{\partial g_i}{\partial x_2} \Big|_{(1-t)a+tb} (b_2 - a_2),$$

y por el teorema del valor medio existe un t_i tal que

$$f_i(b) - f_i(a) = h_i(1) - h_i(0) = h'_i(t_i)(1 - 0),$$

luego

$$|f_i(b) - f_i(a)| = |h'_i(t_i)| \leq K|b_1 - a_1| + K|b_2 - a_2| \leq 2K\|b - a\|,$$

de donde a su vez

$$\|f(b) - f(a)\|^2 = (f_1(b) - f_1(a))^2 + (f_2(b) - f_2(a))^2 \leq 8K^2\|b - a\|^2 \leq 9K^2\|b - a\|^2,$$

luego $\|f(b) - f(a)\| \leq 3K\|b - a\|$. ■

DEMOSTRACIÓN DEL TEOREMA DE LA FUNCIÓN INVERSA: En las condiciones del teorema, tomamos un punto $x_0 \in U$, llamamos $y_0 = f(x_0)$. Basta demostrar que f^{-1} es diferenciable en y_0 . Para ello, previamente demostraremos que f^{-1} es continua en y_0 , es decir, que si $\epsilon > 0$, vamos a encontrar un $\delta > 0$ tal que si $\|y - y_0\| < \delta$, entonces $\|f^{-1}(y) - x_0\| < \epsilon$.

Como U es abierto, podemos suponer que la bola cerrada

$$B = \{x \in \mathbb{R}^2 \mid \|x - x_0\| \leq \epsilon\}$$

está contenida en U . La circunferencia

$$S = \{x \in \mathbb{R}^2 \mid \|x - x_0\| = \epsilon\}$$

es compacta, luego por el teorema de Weierstrass [ITAn 3.31], la función continua $\|f(x) - y_0\|$ toma su valor mínimo en uno de sus puntos, digamos x_1 . Tomamos $\delta = \|f(x_1) - y_0\|/2$ y vamos a ver que cumple lo requerido.

Consideramos, pues, un punto tal que $\|y - y_0\| < \delta$ y aplicamos el teorema de Weierstrass a la bola compacta B y a la función continua $\|f(x) - y\|$, con el que concluimos que existe un $z \in B$ donde $\|f(z) - y\|$ es mínimo. Vamos a probar que $f(z) = y$, así como que $z \notin S$, con lo que tendremos que $\|f^{-1}(y) - x_0\| < \epsilon$, como tenemos que probar.

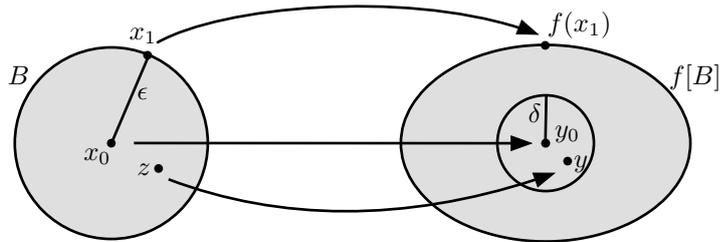
En efecto, por una parte, por la minimalidad de z , tenemos que

$$\|f(z) - y\| \leq \|f(x_0) - y\| < \delta,$$

mientras que si $x \in S$, por la desigualdad triangular

$$\|f(x) - y\| \geq \|f(x) - y_0\| - \|y_0 - y\| > \|f(x_1) - y_0\| - \delta \geq 2\delta - \delta = \delta,$$

donde hemos usado también la minimalidad de x_1 . Esto prueba que $x \neq z$, luego $z \notin S$.



Ahora usamos que si z hace mínima la distancia $\|f(z) - y\|$, también hace mínimo a su cuadrado, es decir, a la función

$$h(x) = (f_1(x) - y_1)^2 + (f_2(x) - y_2)^2.$$

Por consiguiente, las derivadas parciales de h se tienen que anular en z :

$$\left. \frac{\partial h}{\partial x_1} \right|_z = \left. \frac{\partial h}{\partial x_2} \right|_z = 0.$$

Explícitamente,

$$2(f_1(z) - y_1) \frac{\partial f_1}{\partial x_i} \Big|_z + 2(f_2(z) - y_2) \frac{\partial f_2}{\partial x_i} \Big|_z = 0,$$

y esto puede expresarse matricialmente como

$$(f(z) - y) \begin{pmatrix} \frac{\partial f_1}{\partial x_1} \Big|_z & \frac{\partial f_1}{\partial x_2} \Big|_z \\ \frac{\partial f_2}{\partial x_1} \Big|_z & \frac{\partial f_2}{\partial x_2} \Big|_z \end{pmatrix} = (0, 0),$$

pero la matriz es la traspuesta de la matriz jacobiana de f en z , luego tiene su mismo determinante no nulo, luego tiene inversa y, multiplicando por ella, concluimos que $f(z) - y = 0$, como había que probar.

Pasamos ya a probar que f^{-1} es diferenciable en y_0 . En primer lugar probamos que no perdemos generalidad si suponemos que $x_0 = y_0 = (0, 0)$. En efecto, basta considerar la aplicación $g : U^* \rightarrow V^*$ dada por

$$g(x) = f(x + x_0) - y_0,$$

donde

$$U^* = \{x - x_0 \mid x \in U\}, \quad V^* = \{y + y_0 \mid y \in V\}.$$

Como f es de clase C^1 en U , se cumple que g es de clase C^1 en U^* y su inversa es

$$g^{-1}(y) = f^{-1}(y + y_0) - x_0.$$

Se cumple que $(0, 0) \in U^*$, y si probamos que g^{-1} es diferenciable en el punto $f(0, 0) = (0, 0)$, esto implicará que f^{-1} es diferenciable en y_0 , como queremos probar. Además, se cumplen las relaciones

$$\frac{\partial g_i}{\partial x_j} \Big|_x = \frac{\partial f_i}{\partial x_j} \Big|_{x+x_0},$$

de donde se sigue que el determinante de $Jg(x)$ coincide con el de $Jf(x + x_0)$, luego no se anula. Así, cambiando f por g , podemos suponer $x_0 = y_0 = (0, 0)$.

En segundo lugar probamos que no perdemos generalidad si suponemos que la matriz jacobiana $Jf(0, 0)$ es la matriz identidad. En efecto, sea $M = Jf(0, 0)$, que por hipótesis tiene determinante no nulo, luego tiene inversa M^{-1} . Sea $h : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la aplicación lineal de matriz M^{-1} y sea ahora $g : U \rightarrow h^{-1}[V]$ la función compuesta que resulta de aplicar f y luego h . Su inversa g^{-1} es la composición de h^{-1} con f^{-1} .

Por la regla de la cadena sabemos que g es de clase C^1 y que su matriz jacobiana es $Jg(x) = Jf(x) \cdot M^{-1}$, que tiene determinante no nulo, porque el determinante de un producto de matrices es el producto de los determinantes. En particular vemos que $Jg(0, 0) = M \cdot M^{-1} = I$.

Si probamos que g^{-1} es diferenciable en $(0, 0)$, por la regla de la cadena lo mismo le sucederá a la composición de h con g^{-1} , que es f^{-1} . Por lo tanto, cambiando f por g , podemos suponer que $Jf(0, 0) = I$.

Como las derivadas parciales de f son continuas, tenemos que

$$\lim_{x \rightarrow (0,0)} \frac{\partial f_1}{\partial x_1} = \lim_{x \rightarrow (0,0)} \frac{\partial f_2}{\partial x_2} = 1, \quad \lim_{x \rightarrow (0,0)} \frac{\partial f_1}{\partial x_2} = \lim_{x \rightarrow (0,0)} \frac{\partial f_2}{\partial x_1} = 0,$$

de donde se sigue a su vez que existe un $\eta > 0$ tal que si $\|x\| < \eta$ entonces

$$\left| \frac{\partial(f_i - h_i)}{\partial x_j} \right| \leq \frac{1}{6}, \quad i, j = 1, 2,$$

donde h es la aplicación identidad dada por $h(x, y) = (x, y)$.

Como f^{-1} es continua en $(0, 0)$, existe un $\delta > 0$ tal que si $\|v\| < \delta$, entonces $\|f^{-1}(v)\| < \eta$. Llamemos $u = f^{-1}(v)$, de modo que $f(u) = v$. Por el teorema anterior, como $\|u\| < \eta$ (y también $\|(0, 0)\| < \eta$), se cumple que

$$\|(f - h)(u) - (f - h)(0, 0)\| = \|f(u) - u - (0, 0) + (0, 0)\| = \|v - u\| < \frac{1}{2}\|u\|.$$

Por lo tanto,

$$\|u\| = \|u - v + v\| \leq \|u - v\| + \|v\| < \frac{1}{2}\|u\| + \|v\|,$$

luego

$$\|v\| > \frac{1}{2}\|u\|.$$

Así pues:

$$\frac{\|f^{-1}(v) - f^{-1}(0, 0) - v \cdot I\|}{\|v\|} < 2 \frac{\|u - f(u)\|}{\|u\|} = 2 \frac{\|f(u) - f(0, 0) - u \cdot I\|}{\|u\|},$$

y la última expresión tiende a 0 porque f es diferenciable en $(0, 0)$ con matriz jacobiana I , luego la primera expresión también tiende a 0, lo que significa que existe

$$\lim_{v \rightarrow (0,0)} \frac{\|f^{-1}(v) - f^{-1}(0, 0) - v \cdot I\|}{\|v\|} = 0,$$

luego f^{-1} es diferenciable en $(0, 0)$ con matriz jacobiana I . ■

5.4 La energía mecánica

Como primera aplicación notable del cálculo diferencial de varias variables que hemos presentado vamos a discutir un concepto físico fundamental: la energía. En realidad el concepto de energía es muy amplio y aquí sólo vamos a tratar los casos particulares más sencillos. En general, como primera aproximación, podemos decir que la energía es la capacidad de actuar, que algo tiene energía cuando tiene capacidad de generar movimiento, o de romper algo, o de calentarlo, etc. Aquí vamos a tratar únicamente con las formas de energía más directamente relacionadas con el movimiento.

Energía cinética y trabajo Si el lector se pone una piedra en la cabeza, ésta no le hará daño alguno. Si la levanta un centímetro y la deja caer sobre su cabeza, tampoco pasará nada grave, pero si la piedra se dirige hacia su cabeza a gran velocidad, hará bien en apartarse. No obstante, si la “piedra” es un granito de arena, aunque se dirija hacia su cabeza a gran velocidad, tampoco es algo grave.

En general, salvo que tenga otras características, como ser radiactivo, estar muy caliente, etc., un cuerpo en reposo es “inofensivo”, mientras que un cuerpo en movimiento será más peligroso cuanto más masa tenga y más rápidamente se mueva. La dirección del movimiento es irrelevante: una bala que se dirige en una dirección es igual de mortífera que otra de la misma masa que se dirija a la misma velocidad en otra dirección. Tampoco hay diferencia entre ser golpeado por una piedra que se mueve en línea recta que por otra que gira atada a una cuerda a la misma velocidad. Lo que importa es el módulo de su velocidad.

Cada cuerpo, por el mero hecho de estar en movimiento, posee una energía, una capacidad de poner en movimiento a otros cuerpos mediante un choque, o de romperlos, etc. Dicha energía es lo que se conoce como energía cinética, y las observaciones precedentes apuntan a que depende tanto de su masa como de su velocidad. Podría pensarse que una forma de cuantificar la energía cinética de un móvil sería definirla como $E_c = mv$, pero vamos a ver que la definición que “funciona correctamente” es ésta:

Definición 5.21 Se define la *energía cinética* de un objeto de masa m que se mueve a una velocidad v como

$$E_c = \frac{1}{2}mv^2.$$

Observamos que si medimos la masa en kg y la velocidad en m/s, la energía se mide en $\text{kg m}^2/\text{s}^2 = (\text{kg m}/\text{s}^2)\text{m} = \text{N m}$.

Vamos a explicar por qué es razonable esta definición de energía cinética y a la vez entenderemos por qué es razonable pensar que sus unidades son newtons multiplicados por metros en lugar de kilogramos por metro cuadrado entre segundos al cuadrado. Para ello tenemos que introducir la noción de trabajo:

Supongamos que un objeto puntual se mueve siguiendo una trayectoria $\vec{r}(t)$ y que $\vec{F}(t)$ una fuerza que actúa sobre él en cada instante (no necesariamente la fuerza total). Definimos el *trabajo* realizado por dicha fuerza en un intervalo de tiempo $[t_0, t_1]$ como

$$W = \int_{t_0}^{t_1} \vec{F} \cdot d\vec{r} = \int_{t_0}^{t_1} \vec{F} \cdot \frac{d\vec{r}}{dt} dt = \int_{t_0}^{t_1} \vec{F} \cdot \vec{v} dt.$$

Podemos considerar la primera integral como una mera notación que se refiere a cualquiera de las siguientes.

Recordemos de la sección 4.1 que la aceleración se descompone en una aceleración normal \vec{a}_N ortogonal a \vec{v} y una aceleración tangencial a_τ , donde $\tau = \vec{v}/v$

y $a = dv/dt$. Por lo tanto, si $\vec{F} = m\vec{a}$ es la fuerza total que actúa sobre la partícula, tenemos la descomposición

$$\vec{F} = ma\vec{\tau} + m\vec{a}_N,$$

y así, como $\vec{a}_N \cdot \vec{v} = 0$ y $\vec{\tau} \cdot \vec{v} = v$, el trabajo que realiza es

$$W = \int_{t_0}^{t_1} ma\vec{\tau} \cdot \vec{v} dt = \int_{t_0}^{t_1} mav dt = \int_{t_0}^{t_1} \frac{d}{dt} \left(\frac{1}{2}mv^2 \right) dt = \left[\frac{1}{2}mv^2 \right]_{t_0}^{t_1} = \Delta E_c,$$

donde $\Delta E_c = E_c(t_1) - E_c(t_0)$ es el incremento de energía cinética que se produce en el intervalo de tiempo considerado.

Acabamos de probar lo siguiente:

El trabajo realizado por la fuerza total que actúa sobre una partícula en un intervalo de tiempo es igual a la variación de su energía cinética en dicho intervalo.

El trabajo realizado por una fuerza que actúa sobre un objeto sobre el que actúan más fuerzas no tiene una interpretación directa, pero observemos que el trabajo realizado por una suma de fuerzas es claramente la suma de los trabajos que realiza cada una de ellas, luego el resultado precedente puede reformularse diciendo que la variación de la energía cinética de un cuerpo es igual a la suma de los trabajos realizados por cada una de las fuerzas que actúan sobre él.

Observemos que si la fuerza es constante y la trayectoria es rectilínea, entonces

$$W = \int_{t_0}^{t_1} \vec{F} \cdot \vec{v} dt = \pm \int_{t_0}^{t_1} Fv dt = \pm F \int_{t_0}^{t_1} v dt = \pm Fs,$$

donde s es el espacio recorrido por la partícula en el intervalo de tiempo considerado y el signo es positivo si la fuerza actúa en el sentido del movimiento o negativo si actúa en sentido opuesto.

Por ello, en el sistema internacional de unidades, el trabajo y la energía se miden en *julios*, donde un julio ($1 \text{ J} = 1 \text{ N}\cdot\text{m}$) es el trabajo que realiza una fuerza constante de 1 N cuando actúa sobre una partícula que se desplaza 1 m en línea recta (o la energía que ésta gana o pierde).

Notemos que el trabajo que realiza una fuerza depende únicamente de su componente tangencial. En particular una fuerza que sea siempre normal a la trayectoria no realiza trabajo alguno. Por eso hay que distinguir “trabajo” de “esfuerzo”. Por ejemplo, si atamos una piedra a una cuerda y la hacemos girar horizontalmente a velocidad constante, realizaremos un trabajo en el lapso de tiempo en que la piedra pasa de estar en reposo a alcanzar una velocidad estable, pero a partir de ese momento, todo nuestro esfuerzo (la fuerza que realizamos) se destina a mantener la piedra en una órbita circular, pero sin modificar su velocidad, por lo que no estamos realizando ningún trabajo sobre ella: su energía cinética permanece constante. Si en un momento dado soltamos la cuerda y la

pedra sale disparada en línea recta, su energía cinética seguirá siendo la misma, y la piedra es igual de “peligrosa” cuando gira gracias a nuestro esfuerzo que cuando se mueve en línea recta a la misma velocidad. Nuestro esfuerzo por mantenerla girando no contribuye a hacerla más “peligrosa” (no le comunica más energía).

Energía potencial y energía mecánica Supongamos que ponemos una piedra de 10 kg en un montacargas y la elevamos hasta una altura de 30 m. Sobre la piedra actúan dos fuerzas: su peso y la reacción del suelo del montacargas, en sentido opuesto. Cuando el montacargas empieza a moverse, durante un breve instante, ejerce sobre la piedra una fuerza mayor que su peso, que le comunica una pequeña energía cinética, pero dicha fuerza enseguida se iguala a la del peso, porque la piedra sube a velocidad constante. Esto significa que, durante el ascenso, el trabajo positivo que realiza el montacargas sobre la piedra se compensa con el trabajo negativo que hace la gravedad terrestre. Dicho trabajo es fácil de calcular: La fuerza ejercida es $F = 9.8 \cdot 10 \text{ N}$ (de la misma intensidad que el peso), luego el trabajo es $W = 98 \cdot 30 = 2940 \text{ J}$. La gravedad realiza el mismo trabajo, pero negativo (en dirección opuesta al movimiento). Cuando termina el trayecto, la gravedad se encarga de restarle a la piedra la energía cinética que había adquirido, y se queda de nuevo en reposo.

Podría pensarse que el trabajo realizado por el montacargas “se ha perdido”, pues ha sido cancelado por la fuerza gravitatoria, y la piedra no ha ganado energía durante el ascenso, pero eso se puede matizar. Vamos a ver que, en cierto sentido, esos 2940 J de trabajo realizado por el montacargas “se han almacenado” en la piedra, y para “recuperarlos” sólo tenemos que dejarla caer. Si dejamos caer una piedra desde una altura de 30 m, sabemos calcular que su velocidad al llegar al suelo será de 24.25 m/s, por lo que su energía cinética (para una masa de 10 kg) es precisamente de 2940 J.

Podemos pensar que la gravedad terrestre le devuelve a la piedra en su caída los 2940 J que le había quitado durante el ascenso, que la Tierra había contraído una “deuda” de ese importe con la piedra, y vamos a ver que ese tipo de “deudas” son una forma de energía, lo que se conoce como “energía potencial”.

Para concretar matemáticamente estas ideas necesitamos una sencilla aplicación de la regla de la cadena:

Teorema 5.22 Si $V : A \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ es diferenciable y $r : [a, b] \rightarrow A$ es derivable, entonces

$$\int_a^b \nabla V \cdot \frac{dr}{dt} dt = V(r(b)) - V(r(a)).$$

DEMOSTRACIÓN: Basta observar que, por la regla de la cadena,

$$\frac{dV(r(t))}{dt} = \frac{\partial V}{\partial x} \frac{dx}{dt} + \frac{\partial V}{\partial y} \frac{dy}{dt} + \frac{\partial V}{\partial z} \frac{dz}{dt} = \nabla V \cdot \frac{dr}{dt},$$

luego $V(r(t))$ es una primitiva del integrando. ■

Consideremos ahora la gravitación que produce una estrella o planeta de masa M . Si está en una posición \vec{r}_0 , entonces cada partícula de masa m situada en una posición \vec{r} experimenta una atracción dada por

$$\vec{F} = -\frac{GMm}{\|\vec{r} - \vec{r}_0\|^3}(\vec{r} - \vec{r}_0).$$

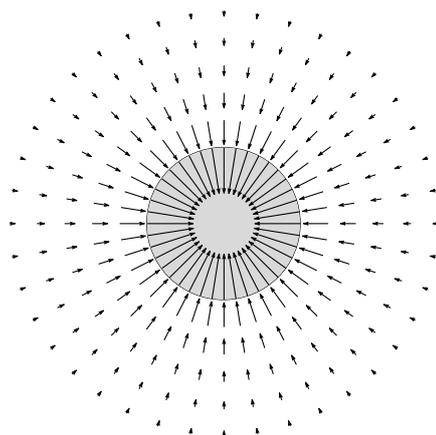
Una aplicación que a cada punto del espacio le asigna un vector se llama *campo vectorial*, y si esos vectores representan una fuerza tenemos un *campo de fuerzas*. Así, podemos decir que una estrella o un planeta genera a su alrededor un campo de fuerzas. En realidad esto no es exacto porque \vec{F} no depende sólo de la estrella o el planeta, sino también de la masa m del objeto que experimenta la fuerza, por lo que es mejor considerar que el campo gravitatorio generado por un cuerpo de masa M es

$$\vec{E} = -\frac{GM}{\|\vec{r} - \vec{r}_0\|^3}(\vec{r} - \vec{r}_0),$$

de modo que cada cuerpo de masa m sometido a dicho campo experimenta una fuerza $\vec{F} = m\vec{E}$.

Por supuesto, todo cuerpo (aunque no sea una estrella o un planeta) genera un campo gravitatorio, pero aquí vamos a considerar campos que no varían en el tiempo, lo cual requiere que podamos suponer que el cuerpo de masa M permanece inmóvil en un sistema de referencia inercial, lo cual a su vez supone despreciar la fuerza que sobre él ejerce el otro cuerpo de masa m o, equivalentemente, despreciar la fuerza de inercia que habría que considerar en un sistema de referencia en dichas condiciones, y esto es razonable si M es mucho mayor que m .

Por ejemplo, la figura siguiente muestra el campo gravitatorio \vec{E} generado por un cuerpo. Si suponemos que es la Tierra o el Sol, hay que tener en cuenta que las flechas no están a escala, pues serían microscópicas:



Ahora observamos que $\vec{E} = -\nabla V$, donde V es la función dada por

$$V = -\frac{GM}{\|\vec{r} - \vec{r}_0\|}.$$

En efecto, $V = -GM((x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2)^{-1/2}$, luego

$$\frac{\partial V}{\partial x} = \frac{1}{2}GM((x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2)^{-3/2}2(x - x_0) = \frac{GM}{\|\vec{r} - \vec{r}_0\|^3}(x - x_0),$$

e igualmente con las otras dos variables, por lo que

$$-\nabla V = -\frac{GM}{\|\vec{r} - \vec{r}_0\|^3}(\vec{r} - \vec{r}_0) = \vec{E}.$$

Por consiguiente, la fuerza gravitatoria que experimenta un cuerpo de masa m es $\vec{F} = -\nabla(mV)$.

Definición 5.23 Cuando un campo vectorial es de la forma $-\nabla V$, para cierta función V , se dice que el campo es *conservativo* y que la función V es una *función potencial* del campo. En el caso concreto del campo de fuerzas que actúa sobre un cuerpo, la función V se llama *energía potencial* del cuerpo (relativa a dicho campo).

El interés de todo esto estriba en que, por el teorema anterior, el trabajo que realiza una fuerza conservativa sobre un cuerpo que se mueve siguiendo una trayectoria $\vec{r}(t)$ es

$$W = \int_{t_0}^{t_1} \vec{F} \cdot d\vec{r} = -V(\vec{r}(t_1)) + V(\vec{r}(t_0)) = -\Delta V,$$

donde $\Delta V = V(t_1) - V(t_0)$. Así pues:

El trabajo que realiza una fuerza conservativa sobre un cuerpo en un intervalo de tiempo es igual a la disminución de su energía potencial en dicho intervalo (si el trabajo es positivo, la energía potencial disminuye, y viceversa).

Observemos ahora que la suma de campos conservativos es de nuevo un campo conservativo, pues

$$\vec{F}_1 + \vec{F}_2 = -\nabla V_1 - \nabla V_2 = -\nabla(V_1 + V_2),$$

luego si llamamos \vec{F}_c a la suma de todas las fuerzas que actúan sobre un objeto determinadas por campos conservativos, \vec{F}_{nc} a la suma de todas las fuerzas restantes que actúan sobre el cuerpo y llamamos W_c y W_{nc} a los trabajos realizados por ellas, respectivamente, entonces el trabajo de la fuerza total que actúa sobre el cuerpo es

$$\Delta E_c = W = W_c + W_{nc} = -\Delta E_p + W_{nc},$$

luego $\Delta E_m = W_{nc}$, donde la *energía mecánica* del cuerpo se define como la suma $E_m = E_c + E_p$ de su energía cinética y su energía potencial, y la conclusión es el *principio de conservación de la energía mecánica*:

La variación de la energía mecánica de un cuerpo en un intervalo de tiempo es igual al trabajo que realizan sobre él en dicho intervalo las fuerzas no conservativas. En particular, si sobre un cuerpo sólo actúan fuerzas conservativas, su energía mecánica permanece constante.

En particular, hemos visto que la energía potencial gravitatoria de un cuerpo de masa m sometido a la atracción de otro cuerpo masivo de masa M viene dada por $E_p = mV$, donde el potencial gravitatorio es

$$V = -\frac{GM}{r},$$

siendo r la distancia entre ambas partículas.

Hay que tener presente que una función potencial para un campo está determinada únicamente por la condición $\vec{F} = -\nabla V$, de donde resulta que si V es una función potencial, cualquier función de la forma $V + c$, donde c es una constante, lo es también. Esto significa que no podemos hablar de la energía potencial que tiene “de verdad” un cuerpo, sino que su valor numérico depende de la elección arbitraria de la constante c , pero lo que sí es objetivo es la variación de la energía potencial de un cuerpo en un intervalo de tiempo determinado.

Por ejemplo, en el caso de un campo gravitatorio es habitual tomar como V precisamente la función que acabamos de indicar, pero si consideramos la aproximación de la fuerza gravitatoria en la superficie terrestre, dada por el campo constante $\vec{F} = (0, 0, -mg)$, que corresponde al campo gravitatorio $\vec{E} = (0, 0, -g)$, éste también deriva de un potencial, a saber, $V = gz$, por lo que la energía potencial gravitatoria de un cuerpo es

$$E_p = mgh,$$

donde $h = z$ es su “altura” medida desde donde hayamos querido tomar el origen del sistema de referencia.

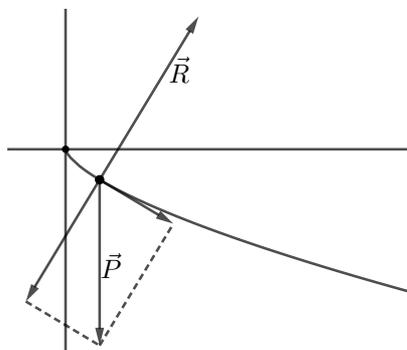
Por ejemplo, antes hemos visto que si subimos una piedra de 10 kg por un montacargas a velocidad constante, el trabajo ejercido sobre la piedra (y compensado por la gravedad) es $W = 2940$ J, pero esto es $W = mgh$, el aumento de energía potencial de la piedra. Ahora sabemos que el resultado sería el mismo aunque el montacargas no fuera a velocidad constante, aunque hubiera acelerado y frenado varias veces, pues el trabajo realizado sobre la piedra sería igual a su aumento de energía mecánica. Como la energía cinética era nula al principio y al final, el trabajo realizado por el montacargas sobre la piedra tiene que ser igual en cualquier caso —independientemente de las condiciones en que se produzca el ascenso— a la energía potencial adquirida por la piedra, mgh .

Similarmente, cuando la piedra cae y llega al suelo $h = 0$, puesto que sobre ella no ha actuado ninguna fuerza no conservativa, su energía mecánica al final tiene que ser la misma que al principio, y como al final su energía potencial se ha reducido a 0, su energía cinética tiene que ser

$$E_c = \frac{1}{2}mv^2 = mgh,$$

luego la velocidad final será $v = \sqrt{2gh}$. A esta conclusión podemos llegar igualmente planteando la ecuación de la trayectoria, calculando el tiempo que tarda la piedra en llegar al suelo y a su vez la velocidad que habrá adquirido al cabo de este tiempo, pero el cálculo a través de la energía es mucho más simple y, además, nos permite concluir de nuevo que la velocidad será la misma tanto si cae en vertical como si cae por un tobogán, incluso si éste tiene subidas y bajadas como una montaña rusa (en la medida en que podamos despreciar el rozamiento, que sería una fuerza no conservativa).

En efecto, si la piedra se desliza sin rozamiento por un tobogán, sobre ella actúan dos fuerzas, su peso \vec{P} y la reacción \vec{R} de la superficie, que es ortogonal a la misma y cuyo único efecto es compensar la componente normal y proporcionar la aceleración normal necesaria para que la trayectoria se curve siguiendo el tobogán, pero, como es una fuerza ortogonal a la trayectoria, no efectúa ningún trabajo sobre la piedra, luego no modifica su energía mecánica.



Por consiguiente, la energía cinética cuando la piedra llega a la base del tobogán coincide con la energía potencial que tenía al principio, $E = mgh$, luego la velocidad final sigue siendo $v = \sqrt{2gh}$, la misma que si la piedra hubiera caído verticalmente. En particular, la curva concreta que describa el tobogán es irrelevante mientras la altura descendida sea la misma.

Ejemplo: La curva isócrona de Leibniz En 1687 Leibniz planteó el problema siguiente, que fue resuelto por Jacques Bernoulli en 1690:

Encontrar la forma de una rampa que haga que los cuerpos que caen por ella se muevan con velocidad vertical constante.

Queremos una trayectoria $r(t) = (x(t), y(t))$ tal que $v_0 = -y'$ sea constante, con lo que $y = -v_0 t$, si tomamos $(0, 0)$ como posición inicial. Vamos a suponer además que $x'(0) = 0$.⁵ La conservación de la energía mecánica nos da entonces

⁵En realidad esto no supone ninguna pérdida de generalidad: podríamos resolver el problema sin esta hipótesis y veríamos que la trayectoria puede prolongarse hacia atrás en el tiempo hasta un punto en el que la velocidad horizontal es nula, por lo que podemos tomar ese instante como punto de partida, pero el cálculo en general queda más farragoso.

que

$$\frac{1}{2}mv_0^2 = \frac{1}{2}m(x'^2 + v_0^2) - mgv_0t,$$

de donde se sigue que

$$x' = \sqrt{2gv_0t},$$

luego

$$x = \int_0^t \sqrt{2gv_0\tau} d\tau = \frac{2}{3}\sqrt{2gv_0t^3}.$$

Por consiguiente, la curva buscada es

$$r(t) = \left(\frac{2}{3}\sqrt{2gv_0t^3}, -v_0t\right).$$

Si queremos únicamente la descripción geométrica de la trayectoria, no la parametrización temporal, observamos que se trata de la curva formada por los puntos (x, y) que cumplen la ecuación

$$x^2 = -a^2y^3,$$

donde

$$a = \frac{2}{3} \frac{\sqrt{2g}}{v_0}.$$

En general, las curvas de la forma

$$y^2 = a^2x^3$$

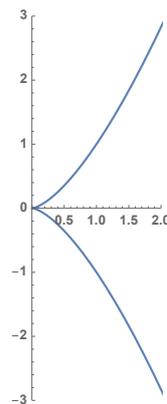
se llaman *parábolas semicúbicas*, con lo que podemos concluir que las curvas isócronas de Leibniz son las parábolas semicúbicas “tumbadas”, pero hay que tener presente que la curva $x^2 - a^2y^3$ sólo es isócrona para objetos que lleguen a su cúspide con velocidad $(0, -v_0)$, donde

$$v_0 = \frac{2}{3} \frac{\sqrt{2g}}{a}.$$

En otras palabras: cada velocidad v_0 requiere su propia curva isócrona. La rampa que muestra la figura precedente es precisamente la isócrona correspondiente a una velocidad de 1 m/s. ■

Volviendo al caso general de una partícula que cae por un tobogán que cae por una rampa, veamos cómo calcular el tiempo que tarda en caer. Si el tobogán está determinado por la función $y(x)$ y llamamos $y(s)$ a la altura en función de la distancia a un punto prefijado medida sobre la curva, e y_0 a la altura inicial (donde la partícula tenía velocidad nula), entonces, cuando se encuentra a una altura $y(s)$, su energía mecánica es

$$E = \frac{1}{2}mv^2 + mgy(s),$$



y tiene que coincidir con la energía mecánica inicial, mgy_0 , luego

$$\frac{1}{2}mv^2 = mg(y_0 - y(s)),$$

de donde

$$\frac{ds}{dt} = v = \pm\sqrt{2g(y_0 - y(s))}, \quad (5.6)$$

donde el signo depende de si s aumenta o disminuye con el tiempo. Esto nos da una ecuación diferencial con variables separables:

$$dt = \pm \frac{1}{\sqrt{2g}} \frac{ds}{\sqrt{y_0 - y(s)}},$$

que al integrarla se convierte en

$$T = \int_0^T dt = \pm \frac{1}{\sqrt{2g}} \int_{s_0}^{s^T} \frac{ds}{\sqrt{y_0 - y(s)}},$$

En particular, el tiempo T que tarda la partícula en llegar a la base del tobogán cuando se deja caer desde una distancia s_0 a la base (medida sobre el tobogán) es:

$$T = \frac{1}{\sqrt{2g}} \int_0^{s_0} \frac{ds}{\sqrt{y_0 - y(s)}}. \quad (5.7)$$

Por ejemplo, si el tobogán es rectilíneo (es un plano inclinado) y va desde $(0, 0)$ hasta $(d, -h)$, con $l = \sqrt{d^2 + h^2}$, tenemos

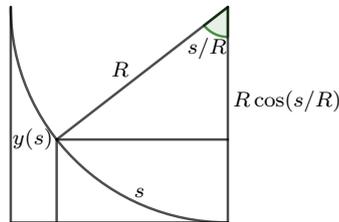
$$y(s) = \frac{h}{l}s,$$

donde medimos la distancia recorrida desde el punto de partida (con lo que $y_0 = 0$). Así:

$$T = \frac{1}{\sqrt{2g}} \int_0^l \frac{ds}{\sqrt{hs/l}} = \sqrt{\frac{l}{2gh}} \int_0^l \frac{ds}{\sqrt{s}} = \sqrt{\frac{l}{2gh}} 2\sqrt{l} = \sqrt{\frac{2}{gh}} l,$$

que es la misma fórmula (4.2) que obtuvimos en el capítulo anterior.

Ejemplo Vamos a calcular el tiempo que tarda en llegar al suelo un objeto que se desliza por un cuarto de circunferencia de radio R :



Si s es la distancia al punto más bajo sobre el arco, entonces el ángulo que lo abarca es s/R , luego la altura es $y(s) = R - R \cos(s/R)$, y la altura máxima es $y_0 = R$, luego la fórmula anterior se particulariza a

$$T = \frac{1}{\sqrt{2g}} \int_0^{\pi R/2} \frac{ds}{\sqrt{R \cos(s/R)}} = \sqrt{\frac{R}{2g}} \int_0^{\pi/2} \frac{dx}{\sqrt{\cos x}}.$$

La integral la calcularemos más adelante (teorema 6.4), y el resultado para $R = 100$ es $T = 5.92$ s.

Vemos cómo el este tiempo es menor que los que calculamos en el ejemplo sobre planos inclinados en la página 225, pero, como indicamos allí, no es el menor posible. Dedicaremos la sección 5.5 a estudiar este problema. ■

Ejemplo: La curva tautócrona Un caso de caída especialmente destacado se da cuando el tobogán tiene forma de cicloide. Hemos visto (página 137) que, para una cicloide de radio R , se cumple que $y(s) = s^2/8R$. Podemos sustituir esta función en la fórmula para T y la integral se puede resolver con la regla del arco coseno, pero en realidad es más sencillo si vamos directamente a la ecuación

$$v^2 = 2g(y_0 - y(s)) = \frac{g}{4R}(s_0^2 - s^2),$$

y derivamos respecto de s , teniendo en cuenta que $v = s'$. Entonces:

$$2s's'' = -\frac{g}{2R}ss',$$

de donde llegamos a la ecuación diferencial

$$s'' = -\frac{g}{4R}s,$$

cuya solución es

$$s = s_0 \cos\left(\frac{1}{2}\sqrt{\frac{g}{R}}t\right), \quad \text{o} \quad y = y_0 \cos^2\left(\frac{1}{2}\sqrt{\frac{g}{R}}t\right).$$

Así, s o y toman el valor 0 en el instante en que

$$\frac{1}{2}\sqrt{\frac{g}{R}}T = \frac{\pi}{2},$$

es decir, en el instante

$$T = \pi\sqrt{\frac{R}{g}},$$

que es independiente de s_0 , es decir, que si situamos un objeto en un tobogán en forma de cicloide de modo que se caiga sin rozamiento, el tiempo que tardará

en llegar a la base no depende de la altura a la que lo hayamos situado.⁶ Puede probarse que la cicloide es la única curva con esta propiedad, y por ello se dice que es la *curva tautócrona* (gr. “el mismo tiempo”).

Naturalmente, si se deja caer una bola por una cicloide y no hay rozamiento, la bola no se detendrá en la base, sino que continuará ascendiendo por la cicloide hasta detenerse a la misma altura de la que había partido en el extremo opuesto, y desde allí volverá a caer, de modo que oscilará con un periodo $4\pi\sqrt{R/g}$ (el tiempo en que tarda en volver al punto inicial), que es independiente del punto de partida.

Esto lo descubrió Christiaan Huygens en 1659. Como hemos explicado en la introducción, Huygens observó que el movimiento de un péndulo es el mismo que el de una bola que oscila sobre un arco de circunferencia, y sabía que el hecho de que el periodo de oscilación no depende de la amplitud es sólo una aproximación válida para amplitudes pequeñas, así que se planteó el encontrar una curva que no fuera un arco de circunferencia para la que esta propiedad se cumpliera de forma exacta. Aunque su razonamiento fue geométrico, no analítico, en términos analíticos Huygens buscaba una curva en la que, al dejar oscilar sobre ella una bola, se cumpliera la ecuación $s'' = -ks$. Ahora bien, si derivamos respecto al tiempo t la relación $v^2 = 2g(y_0 - y)$, obtenemos

$$2s's'' = -2g\frac{dy}{ds}s',$$

y al imponer la relación $s'' = -ks$, resulta

$$\frac{dy}{ds} = \frac{k}{g}s,$$

y, si medimos s desde la base, de modo que $y(0) = s(0) = 0$, esto equivale a

$$y = \frac{k}{2g}s^2,$$

y en el apartado sobre la cicloide en la página 137 vimos cómo a partir de esta condición puede probarse que la curva es una cicloide. Ahora bien, la solución que hemos dado es incompleta, porque hemos empezado buscando una curva que cumpla $s'' = -ks$. Ahora vamos a probar que las cicloides son las únicas curvas tautócronas sin suponer dicha condición.⁷

⁶Si el lector quiere programar la caída de una bola por una cicloide, sólo tiene que igualar

$$y = R(1 - \cos \theta) = y_0 \cos^2 \left(\frac{1}{2} \sqrt{\frac{g}{R}} t \right)$$

y despejar $\theta(t)$, de donde $x(t) = R(\theta(t) + \sin \theta(t))$ y así tenemos la posición $\vec{r}(t) = (x(t), y(t))$ en función del tiempo.

⁷Aquí es importante incidir en que la tautocronía, para una curva en forma de U , se define como que el tiempo de llegada a la base de la curva —no el tiempo de regreso al punto inicial— sea independiente del punto de partida, porque curvas para las que el tiempo de regreso es siempre el mismo hay infinitas más, aparte de la cicloide, pero en ninguna de las otras la bola tarda el mismo tiempo en llegar a la base. Son curvas asimétricas, en las que la rama de bajada no se parece a la de subida que tiene enfrente.

Para ello partimos de la ecuación (5.7):

$$\frac{1}{\sqrt{2g}} \int_0^{s_0} \frac{ds}{\sqrt{y_0 - y(s)}} = T,$$

donde $y(s)$ es la altura de una curva $(x(s), y(s))$ parametrizada por el arco desde su punto más bajo $y(0) = 0$. El miembro izquierdo proporciona entonces el tiempo que tarda un objeto en caer hasta la base desde un punto a una altura $y_0 = y(s_0)$ y T es un tiempo constante. Puesto que el objeto desciende en todo momento, la función $y(s)$ admite una inversa $s(y)$, y podemos aplicar a la integral el cambio de variable $s = s(y)$, con lo que la ecuación se transforma en

$$\frac{1}{\sqrt{2g}} \int_0^{y_0} \frac{1}{\sqrt{y_0 - y}} \frac{ds}{dy} dy = T,$$

donde ahora la incógnita es la función $s(y)$. Si llamamos $f(y) = ds/dy$, la integral es la convolución de las funciones $y^{-1/2}$ y $f(y)$, luego aplicando la transformada de Laplace a ambos miembros resulta

$$\frac{1}{\sqrt{2g}} \sqrt{\frac{\pi}{u}} F(u) = \frac{T}{u},$$

donde hemos usado que $\mathcal{L}\{y^{-1/2}\}(u) = \sqrt{\pi/u}$ y donde llamamos $F(u)$ a la transformada de $f(y)$. Así hemos transformado una ecuación integral en una ecuación algebraica, cuya solución es

$$F(u) = \sqrt{\frac{2g}{\pi}} T \frac{1}{\sqrt{u}}$$

La transformada inversa es evidente:

$$\frac{ds}{dy} = \sqrt{\frac{2g}{\pi}} T \frac{1}{\sqrt{\pi y}},$$

con lo que tenemos una ecuación diferencial de variables separables, cuya solución es (teniendo en cuenta que $y(0) = 0$):

$$s = \frac{\sqrt{2g}}{\pi} T 2\sqrt{y},$$

o, equivalentemente,

$$y = \frac{\pi^2}{8gT} s^2,$$

con lo que llegamos de nuevo a la caracterización de la cicloide, que es, pues, la única curva tautócrona. ■

Ejemplo: La curva sinodal El problema siguiente fue planteado y resuelto en 1790 por el ingeniero italiano Teodoro Bonati:

Encontrar una curva tal que un cuerpo que caiga sobre ella desde un punto O tarda en llegar a cada punto el mismo tiempo que otro que caiga por un plano inclinado.

La solución se conoce como *curva sinodal*, que en griego es aproximadamente lo mismo que “confluente”. Si llamamos $r(t) = (x(t), y(t))$ a la trayectoria de un cuerpo que se deja caer sobre la curva desde el punto $(0, 0)$, hemos visto que el tiempo que tarda en recorrer una distancia s_0 viene dado por la fórmula (5.7):

$$T = \frac{1}{\sqrt{2g}} \int_0^{s_0} \frac{ds}{\sqrt{-y(s)}}.$$

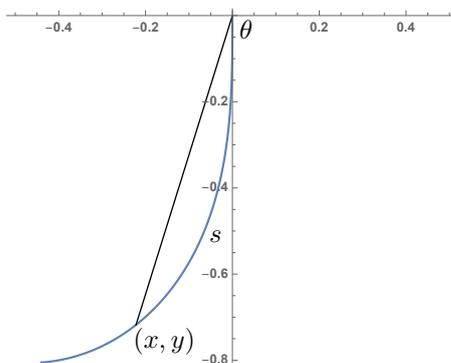
Para el caso de un camino en línea recta que llegue hasta el mismo punto (x, y) , el tiempo viene dado por (4.2), luego queremos que se cumpla la igualdad:

$$\frac{1}{\sqrt{2g}} \int_0^{s_0} \frac{ds}{\sqrt{-y(s)}} = \sqrt{\frac{2}{-gy}} \sqrt{x^2 + y^2},$$

que se simplifica hasta

$$\int_0^{s_0} \frac{ds}{\sqrt{-y(s)}} = 2\sqrt{\frac{x^2 + y^2}{-y}},$$

para todo punto (x, y) de la curva. Vamos a expresar esta igualdad en coordenadas polares. Por conveniencia vamos a suponer también que el tobogán avanza hacia la izquierda, luego el ángulo θ puede variar entre $-\pi/2$ y $-\pi$:



Para ello, en la integral hacemos el cambio de variable $s = s(\theta)$, teniendo en cuenta que, según la fórmula (2.3) —tomando a θ como parámetro, de modo que $\theta' = 1$ — se cumple que

$$\frac{ds}{d\theta} = -\sqrt{\rho^2 + \rho'^2},$$

donde el signo negativo se debe a que, con los criterios que hemos adoptado, s aumenta cuando disminuye θ . Además, $y(s(\theta)) = y(\theta) = \rho(\theta) \operatorname{sen} \theta$. Así pues:

$$-\int_{-\pi/2}^{\theta_0} \frac{\sqrt{\rho^2 + \rho'^2}}{\sqrt{-\rho \operatorname{sen} \theta}} d\theta = 2\sqrt{-\frac{\rho}{\operatorname{sen} \theta_0}}.$$

Derivamos ambos miembros y cambiamos θ_0 por θ :

$$-\frac{\sqrt{\rho^2 + \rho'^2}}{\sqrt{-\rho \operatorname{sen} \theta}} = \frac{1}{\sqrt{-\frac{\rho}{\operatorname{sen} \theta}}} \frac{-\rho' \operatorname{sen} \theta + \rho \cos \theta}{\operatorname{sen}^2 \theta}.$$

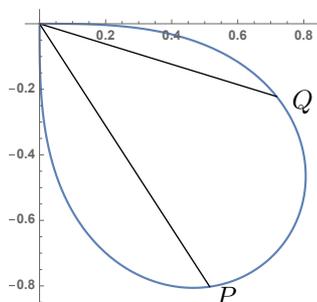
Elevamos al cuadrado:

$$-\frac{\rho^2 + \rho'^2}{\rho \operatorname{sen} \theta} = -\frac{\operatorname{sen} \theta (-\rho' \operatorname{sen} \theta + \rho \cos \theta)^2}{\rho \operatorname{sen}^4 \theta}.$$

y al simplificar la ecuación se reduce a

$$\rho' = \frac{\rho}{\tan \theta},$$

que es la ecuación diferencial (2.10), cuya solución es $\rho = a\sqrt{\operatorname{sen} 2\theta}$, que corresponde a una lemniscata de Bernoulli girada -45° . Por lo tanto, las curvas sinodales son las lemniscatas de Bernoulli (o la mitad de ellas, para ser precisos) en esta posición (o en la simétrica, respecto del eje vertical, obviamente). Por ejemplo, si dejamos caer a la vez tres bolas desde el origen de coordenadas en la figura siguiente, una por la lemniscata y las otras dos por las rectas, la que recorre la lemniscata encontrará a las otras dos en los puntos P y Q .



■

Sistemas binarios Por completitud vamos a ver que la conservación de la energía mecánica es aplicable a un sistema formado por dos partículas puntuales de masas similares, de modo que no podamos despreciar la fuerza que una ejerce sobre la otra.

En general, supongamos que tenemos dos partículas que desde un sistema de referencia inercial tienen posiciones \vec{r}_1 y \vec{r}_2 y que se ejercen mutuamente fuerzas conservativas, $\vec{F}_{12} = -\vec{F}_{21}$ determinadas por funciones potenciales que

dependen únicamente de la distancia a la partícula, es decir, que son de la forma $V_{ij}(\vec{r}) = V(\|\vec{r} - \vec{r}_i\|)$, para cierta función $V(s)$. En el caso de las fuerzas gravitatorias los potenciales son

$$V_{ij}(\vec{r}) = -\frac{Gm_1m_2}{\|\vec{r} - \vec{r}_i\|},$$

con lo que $V(s) = -Gm_1m_2/s$.

El campo de fuerzas asociado al potencial $V_{ij}(\vec{r})$ es

$$\vec{f}_{ij}(\vec{r}) = -\nabla V(\|\vec{r} - \vec{r}_i\|) = -\frac{dV}{ds}(\|\vec{r} - \vec{r}_i\|) \left(\frac{ds}{dx}, \frac{ds}{dy}, \frac{ds}{dz} \right),$$

donde $s = \|\vec{r} - \vec{r}_i\| = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}$, luego

$$\vec{f}_{ij}(\vec{r}) = -\frac{dV}{ds}(\|\vec{r} - \vec{r}_i\|) \left(\frac{x - x_i}{\|\vec{r} - \vec{r}_i\|}, \frac{y - y_i}{\|\vec{r} - \vec{r}_i\|}, \frac{z - z_i}{\|\vec{r} - \vec{r}_i\|} \right) = -\frac{dV}{ds}(\|\vec{r} - \vec{r}_i\|) \frac{\vec{r} - \vec{r}_i}{\|\vec{r} - \vec{r}_i\|},$$

que tiene la dirección (aunque no necesariamente el sentido) de $\vec{r} - \vec{r}_i$.

Supongamos que, sobre la partícula j , además de actuar la fuerza \vec{F}_{ij} que le ejerce la partícula i , se ejerce una fuerza exterior \vec{F}_j . Entonces, la variación de la energía cinética conjunta de ambas partículas en un intervalo de tiempo $[t_0, t_1]$ será

$$\Delta E_c = (W_1 + W_{21}) + (W_2 + W_{12}),$$

donde los dos primeros sumandos son la variación de la energía cinética de la partícula 1 y los otros dos corresponden a la partícula 2. Ahora bien, si llamamos $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$, tenemos que

$$\begin{aligned} W_{12} + W_{21} &= \int_{t_0}^{t_1} -\frac{dV}{ds}(\|\vec{r}_{12}\|) \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|} \frac{d\vec{r}_2}{dt} dt + \int_{t_0}^{t_1} -\frac{dV}{ds}(\|\vec{r}_{21}\|) \frac{\vec{r}_{21}}{\|\vec{r}_{21}\|} \frac{d\vec{r}_1}{dt} dt \\ &= \int_{t_0}^{t_1} -\frac{dV}{ds}(\|\vec{r}_{12}\|) \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|} \frac{d\vec{r}_2}{dt} dt - \int_{t_0}^{t_1} -\frac{dV}{ds}(\|\vec{r}_{12}\|) \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|} \frac{d\vec{r}_1}{dt} dt \\ &= -\int_{t_0}^{t_1} \frac{dV}{ds}(\|\vec{r}_{12}\|) \frac{\vec{r}_{12}}{\|\vec{r}_{12}\|} \frac{d\vec{r}_{12}}{dt} dt = -\int_{t_0}^{t_1} \frac{dV(\|\vec{r}_{12}\|)}{dt} dt = -\Delta E_p, \end{aligned}$$

donde la *energía potencial interna* del sistema formado por las dos partículas es

$$E_p = V(\|\vec{r}_{12}\|).$$

En el caso en que las fuerzas son gravitatorias es

$$E_p = -\frac{Gm_1m_2}{r},$$

donde r es la distancia que separa a las partículas. La conclusión es que si definimos la energía mecánica del sistema de las dos partículas como la suma de la energía cinética y potencial: $E_m = E_c + E_p$, tenemos que $\Delta E_m = W_{\text{ext}}$ donde el *trabajo exterior* es $W_{\text{ext}} = W_1 + W_2$ la suma de los trabajos que las fuerzas externas ejercen sobre las partículas. Por lo tanto, también se cumple en este caso el principio de conservación de la energía mecánica:

La variación en un intervalo de tiempo de la energía mecánica de dos partículas que interactúan mediante fuerzas conservativas es igual al trabajo que en ese intervalo ejercen sobre ellas las fuerzas exteriores. En particular, si no hay fuerzas exteriores (o éstas no ejercen ningún trabajo) la energía mecánica permanece constante.

Notemos que la energía potencial que hemos asignado a un sistema de dos partículas es la misma que habíamos asignado a una de ellas cuando consideramos la otra “fija”, con lo cual, lo único que hemos obtenido es que, al considerar el movimiento de las dos partículas, de modo que ambas contribuyen a la variación de la energía cinética, no debemos sumar las energías potenciales de ambas, sino que la energía potencial que asignábamos a una, es en realidad “de las dos”. ■

La energía y el movimiento planetario Es interesante analizar el movimiento planetario en términos de la energía mecánica. Consideremos un cuerpo de masa m que orbita alrededor de una estrella de masa M mucho mayor situada en el origen de coordenadas.⁸ Descartamos el caso en que el cuerpo se mueve en línea recta o, lo que es lo mismo, suponemos que sigue una trayectoria cónica. Tomando módulos en (4.4) obtenemos que $v^2 = \rho'^2 + \rho^2\omega^2$, luego, recordando que $L = m\rho^2\omega$ es invariante, la energía cinética del cuerpo es

$$E_c = \frac{1}{2}m(\rho'^2 + \rho^2\omega^2) = \frac{1}{2}m(\rho'^2 + \frac{L^2}{m^2\rho^2}),$$

mientras que la energía potencial es

$$V = -\frac{GMm}{\rho}$$

y la energía mecánica es

$$E = \frac{1}{2}m(\rho'^2 + \frac{L^2}{m^2\rho^2}) - \frac{GMm}{\rho} = \frac{1}{2}m\rho'^2 + V_e(\rho),$$

donde

$$V_e(\rho) = \frac{L^2}{2m\rho^2} - \frac{GMm}{\rho}$$

es lo que podemos llamar *energía potencial efectiva*, pues depende únicamente de la posición del cuerpo. Vamos a esbozar su gráfica: es claro que sólo toma el valor 0 en

$$\rho^* = \frac{L^2}{2GMm^2},$$

⁸Según vimos al final de la sección 4.4, si las masas son arbitrarias basta tomar el origen de coordenadas en el centro de masas del sistema y modificar formalmente las masas de ambos objetos para que todas las conclusiones sigan siendo válidas.

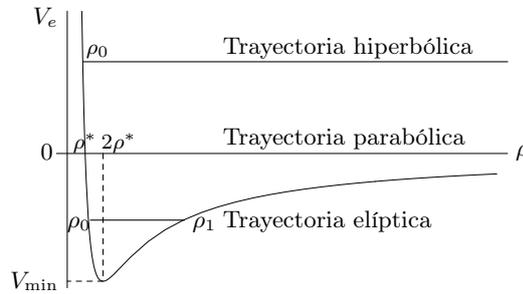
que tiende a $+\infty$ en 0 y que tiende a 0 en $+\infty$. La derivada

$$\frac{dV_e}{dr} = \frac{1}{\rho^2} \left(GMm - \frac{L^2}{m\rho} \right)$$

se anula únicamente en $2\rho^*$, es negativa para $0 < r < 2\rho^*$ (por ejemplo, porque el paréntesis tiende a $-\infty$ en 0) y es positiva para $\rho > 2\rho^*$ (porque el paréntesis tiende a $GMm > 0$ en $+\infty$). Por lo tanto el potencial tiene un mínimo en $2\rho^*$, donde toma el valor

$$V_{\min} = -\frac{G^2 M^2 m^3}{2L^2}.$$

Así pues, la gráfica de V_e tiene el aspecto siguiente:



Los valores ρ^* y $2\rho^*$ tienen una interpretación en virtud de la fórmula (4.8):

$$\rho_0 = \frac{L^2}{GMm^2} (1 + \epsilon)^{-1}$$

que relaciona el perihelio ρ_0 con la excentricidad de la órbita, y que ahora podemos expresar en la forma

$$\rho_0 = \frac{2\rho^*}{1 + \epsilon}.$$

Así, $2\rho^*$ es el radio de la órbita cuando ésta es circular ($\epsilon = 0$) y ρ^* es el perihelio cuando la órbita es parabólica ($\epsilon = 1$). El perihelio de una órbita elíptica cumple $\rho^* \leq \rho_0 < 2\rho^*$ y el de una órbita parabólica cumple $\rho_0 > 2\rho^*$.

De la expresión

$$E = \frac{1}{2} m \rho'^2 + V_e$$

se sigue que la energía mecánica E del cuerpo no puede ser inferior a V_{\min} . A partir de aquí, distinguiamos varios casos:

- Si $E = V_{\min}$, ha de ser necesariamente $\rho' = 0$ y ρ sólo puede tomar el valor ρ^* , luego la trayectoria es circular.

- Si $V_{\min} < E < 0$, entonces los valores posibles para ρ oscilarán entre un mínimo ρ_0 y un máximo ρ_1 , determinados por la ecuación

$$\frac{L^2}{2m\rho^2} - \frac{GMm}{\rho} = E. \quad (5.8)$$

Además ρ' sólo puede valer 0 cuando ρ toma esos valores que hacen $V_e = E$, lo que implica que ρ tiene que oscilar entre estos valores máximo y mínimo (la órbita no puede ser circular con radio constante ρ_0 o ρ_1 porque ya hemos visto que para una órbita circular el radio tiene que ser $2\rho^*$).

- Si $E = 0$, el cuerpo no podrá acercarse a la estrella a menos distancia que ρ^* y ya hemos visto que esto significa que la órbita es una parábola.
- Si $E > 0$ la ecuación (5.8) tiene sólo una raíz positiva $\rho_0 < \rho^*$, que será el perihelio de una trayectoria hiperbólica.

Así pues, la órbita resulta ser elíptica, parabólica o hiperbólica según si la energía mecánica es negativa, nula o positiva. Más precisamente:

- Si la energía total es $E < 0$ se dice que el cuerpo está en un “pozo de potencial”: la energía potencial no puede exceder el valor E , lo que se traduce en que el cuerpo no puede ir más allá de los radios ρ_0 y ρ_1 correspondientes al perihelio y al afelio.
- En cambio, si la energía total es $E \geq 0$, la energía potencial puede tomar valores arbitrariamente próximos a 0, por lo que el cuerpo puede alejarse indefinidamente de la estrella.

En el caso de una trayectoria rectilínea ($L = 0$) la gráfica de

$$V_e(\rho) = -\frac{GMm}{\rho}$$

es una hipérbola, y se razona igualmente que el cuerpo termina cayendo en la estrella si $E < 0$, mientras que si $E \geq 0$, o bien cae sobre ella si se mueve hacia ella o bien escapa si se mueve en dirección opuesta. ■

Otras formas de energía Para terminar esta sección señalaremos que el principio de conservación de la energía mecánica que hemos analizado es sólo un caso particular de una de las leyes físicas más generales y fundamentales: el principio de conservación de la energía.

Cuando una piedra llega al suelo con cierta energía cinética, se para, pero su energía cinética no desaparece, sino que se “disipa”, es decir, se transforma en otras formas de energía: la piedra emite un sonido, lo cual significa que ha puesto en movimiento las moléculas del suelo y las del aire que transmite el sonido, y también habrá calentado inapreciablemente el suelo con el choque, si no es que le ha causado algún desperfecto.

Precisamente, la unidad de energía recibe su nombre en honor al físico James Prescott Joule, que formuló la teoría de que el calor es una forma de energía y obtuvo una primera estimación de su equivalente mecánico. Actualmente se considera por definición que 1 caloría equivale a 4.184 J, donde la caloría se ha definido históricamente de formas distintas no exactamente equivalentes, que esencialmente tienen que ver con la energía necesaria para elevar un grado la temperatura de un gramo de agua, lo cual requiere especificar muchos parámetros.

El calor, la luz, el sonido, la radiación, etc. son formas de energía o transportan energía, de modo que si tenemos en cuenta todas las formas posibles de energía se cumple el *principio de conservación de la energía*, según el cual en todo sistema físico aislado la energía permanece constante, no hay creación ni destrucción de energía, sino, a lo sumo, la energía puede disiparse en formas prácticamente imposibles de medir en condiciones normales.

5.5 El problema de la braquistócrona

Veamos ahora cómo resolvió Johann Bernoulli el problema de la braquistócrona. Para ello partió de la ley de Snell sobre la refracción de la luz. Cuando un rayo de luz pasa de un medio en el que se mueve a velocidad v_1 a otro en el que se mueve a velocidad v_2 , se refracta siguiendo la ley de Snell (1.2):

$$\frac{\operatorname{sen} \alpha_1}{\operatorname{sen} \alpha_2} = \frac{v_1}{v_2},$$

donde α_1 y α_2 son los ángulos que forman con la perpendicular al plano que separa ambos medios el rayo incidente y el rayo refractado, respectivamente. Y Fermat había demostrado que esto se traduce en que el rayo de luz tarda el menor tiempo posible en viajar a las velocidades dadas entre dos puntos cualesquiera de su trayectoria.

Por otro lado, si un cuerpo cae por una rampa descrita por una curva $y(x)$ desde una altura $y(0) = 0$, su energía cinética en un instante t tiene que ser igual al descenso de su energía potencial:

$$\frac{1}{2}mv^2 = -mgy,$$

Despejando:

$$v = \sqrt{-2gy}.$$

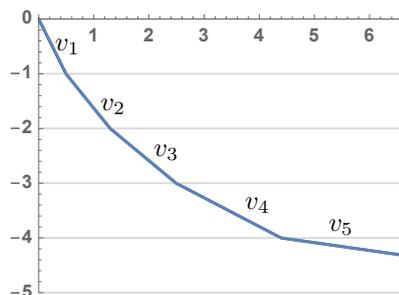
La idea de Bernoulli era que la braquistócrona es la trayectoria que seguiría un rayo de luz que descendiera por un medio cuya densidad fuera decreciendo constantemente, de modo que la velocidad a cada altura $-h$ fuera $\sqrt{2gh}$.

En lugar de seguir el argumento de Bernoulli para ver lo que esto entraña, vamos a tratar de esbozar cómo tiene que ser esa trayectoria. Para ello, vamos a considerar un medio cuya densidad no varíe de forma continua, sino por estratos de espesor h . En la figura siguiente hemos tomado $h = 1$. El primer

tramo representa un rayo de luz que se mueve a velocidad $v_1 = \sqrt{2gh}$ con una inclinación arbitraria θ_1 respecto de la vertical. Si llamamos

$$c = \frac{\text{sen } \theta_1}{v_1},$$

en la figura hemos elegido θ_1 de modo que $c = 0.1$.



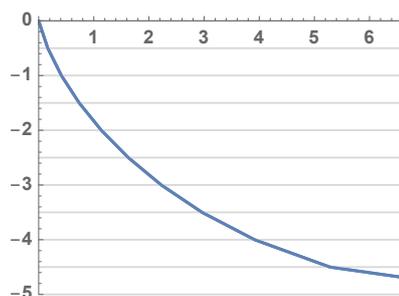
Al llegar al final del primer estrato (a la altura $-h$), la velocidad del rayo de luz pasa a ser $v_2 = \sqrt{2g(2h)}$, y su inclinación varía hasta un ángulo θ_2 según la ley de Snell, de modo que

$$\frac{\text{sen } \theta_2}{v_2} = \frac{\text{sen } \theta_1}{v_1} = c.$$

Y del mismo modo calculamos los tramos siguientes, de modo que $v_i = \sqrt{2gih}$ y

$$\frac{\text{sen } \theta_i}{v_i} = c.$$

Ahora repetimos los mismos cálculos con estratos de espesor $h = 0.5$, manteniendo el mismo valor de c . El resultado es:

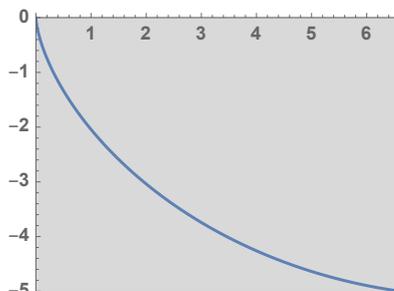


Tenemos una poligonal que en sus vértices $P_i = (x_i, y_i)$ donde $y_i = -ih$ satisface la relación

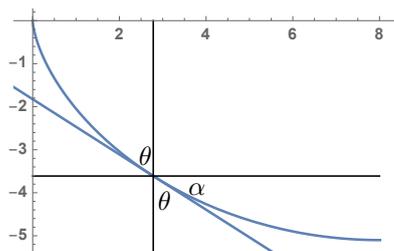
$$\frac{\text{sen } \theta_i}{\sqrt{-2gy_i}} = c,$$

donde θ_i es la pendiente del segmento que llega a P_i (no del que sale de P_i).

La figura siguiente está calculada con $h = 0.01$. Ahora hay representados 500 estratos de espesor 0.01, y la poligonal tiene otros tantos vértices, tan próximos que nada indica en la figura que no se trate de una curva derivable. La pregunta entonces es cuál es esa curva derivable que parece que estamos viendo, ya que no es realmente la poligonal que hemos dibujado.



Nuestra poligonal parece derivable porque la diferencia entre cada ángulo θ_i y θ_{i+1} resulta inapreciable, y ambos se confunden con $\theta(x) = \pi/2 - \alpha(x)$, donde $\alpha(x)$ es la pendiente de la curva derivable que parece que estamos viendo.



Por lo tanto, como $\sin \theta = \cos \alpha$, buscamos una curva $y(x)$ que cumpla

$$\frac{\cos \alpha(x)}{\sqrt{-2gy(x)}} = c.$$

Por otra parte, la pendiente $\alpha(x)$ está relacionada con la derivada por la fórmula $y'(x) = \tan \alpha(x)$ y, como

$$\cos^2 \alpha = \frac{1}{1 + \tan^2 \alpha},$$

la curva que buscamos debe cumplir

$$\frac{1}{\sqrt{-2g(1 + y'^2)y}} = c,$$

o también

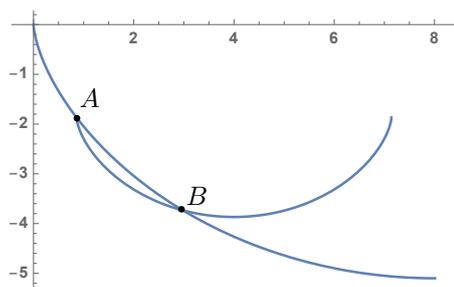
$$-y(1 + y'^2) = \frac{1}{2gc^2} = C, \quad y(0) = 0,$$

y así llegamos a una ecuación diferencial que debe cumplir la curva que buscamos (en nuestro ejemplo, $C = 1/(2g \cdot 0.1^2) = 5.01$).

Bernoulli no llegó a esta ecuación considerando aproximaciones con poligonales, sino mediante razonamientos con infinitésimos, pero en esencia es lo mismo. La solución la hemos calculado ya, pues vemos que la función $-y(x)$ satisface la ecuación diferencial que resolvimos en el ejemplo de la página 135, luego concluimos que la braquistócrona, al igual que la tautócrona, no es sino la cicloide. Concretamente, en nuestro ejemplo, la cicloide de radio $R = c/2 = 2.55$ (que es la que hemos representado en la figura precedente).

Ahora, bien, ¿realmente hemos demostrado que la cicloide es la braquistócrona? Podemos considerar demostrado que la cicloide es la trayectoria que seguiría un rayo de luz que atravesara un medio cuya densidad decrece proporcionalmente a la raíz cuadrada de la profundidad, pero, ¿realmente dicha trayectoria minimiza el tiempo empleado para ir de un punto a otro? Tenemos probado que esto es así cuando la luz pasa de un medio homogéneo a otro también homogéneo, pero de distinta densidad, pero ¿sigue siendo cierto cuando consideramos varios estratos y luego “pasamos al límite”?

Pero podemos formular incluso una objeción mucho más contundente: si el argumento que hemos dado fuera concluyente, podríamos afirmar que la trayectoria por la que un cuerpo tarda menos en caer desde un punto A hasta un punto B de una cicloide es la propia cicloide, pero esto no es cierto:



La trayectoria que minimiza el tiempo en pasar de A a B es la cicloide que tiene en A su punto más alto. La cicloide sólo es braquistócrona cuando se deja caer un objeto en reposo desde su punto más alto, pero no desde un punto intermedio.

Tal y como observó Caratheodory en 1904, a las demás pruebas presentadas se les puede plantear objeciones similares. Sin embargo, en 1718 Johann Bernoulli publicó otra solución del problema de la braquistócrona que no había publicado en 1697 “por razones que ya no estaban vigentes en 1718”, y puede considerarse que es la primera demostración concluyente de que la braquistócrona es la cicloide.

En lugar de analizar este argumento, plagado de infinitésimos, vamos a dar una prueba rigurosa reciente, publicada en 2010 por G. Brookfield, que es completamente elemental y, de hecho, ya hemos realizado buena parte del trabajo que requiere al analizar las coordenadas cicloidales en la página 310.

Ante todo, el teorema 5.17 prueba que si queremos dejar caer un objeto desde un punto A hasta otro B situado a una altura menor (o igual), siempre

existe una única cicloide que empieza en A y pasa por B . Vamos a demostrar que dicha cicloide es la braquistócrona que conecta ambos puntos. Podemos suponer que $A = (0, 0)$ y que B está a la derecha de A . Hay que descartar el caso en que B está debajo de A , pues entonces la trayectoria más rápida es claramente la vertical.

Consideremos un objeto que se deja caer partiendo del reposo por un tobogán (o un tubo) que va de A hasta B en un tiempo T . Su trayectoria vendrá dada por una curva $\vec{r}(t) = (x(t), y(t))$, con $\vec{r}(0) = A$, $v(0) = 0$ y $r(T) = B$.

No es necesario suponer que la curva \vec{r} sea derivable en todos los puntos del intervalo $[0, T]$, sino que podemos admitir que no lo sea en un número finito de puntos (pero sí que hay que suponer que es continua en $[0, T]$). Esto permite, por ejemplo, que el tobogán tenga forma de poligonal.

En el instante inicial, tanto la energía cinética como la potencial son nulas, luego lo mismo tiene que suceder en cualquier otro instante:

$$\frac{1}{2}mv^2 + mgy = 0,$$

luego

$$v^2 = -2gy.$$

En particular, esto implica que la función v (el módulo de $\vec{r}'(t)$) puede extenderse a una función continua en $[0, T]$ aunque $\vec{r}'(t)$ no esté definida en un número finito de puntos. Más aún, podemos deducir algunas características cualitativas de la curva:

- Es imposible que, en su trayectoria, el objeto pase por un punto de altura $y > 0$, pues para ello tendría que tener energía mecánica positiva.
- Si en algún momento el cuerpo alcanza la altura $y = 0$, lo hará con velocidad nula, luego, o bien estamos hablando del punto A de partida, o bien es el punto B , o bien se trata de un pico en la trayectoria, porque el cuerpo no puede moverse con velocidad nula por un tramo horizontal. Es razonable suponer que esto sucede a lo sumo en un número finito de puntos.
- Si el tubo pasa por puntos situados a la izquierda de A , es decir, con $x < 0$, seguro que la trayectoria no minimiza el tiempo, pues en tal caso podríamos considerar el supremo T_0 del conjunto $\{t \in [0, T] \mid x(t) = 0\}$. Por la continuidad de x , será $x(T_0) = 0$, luego $T_0 < T$, y en el intervalo $]T_0, T]$ se cumplirá $x > 0$. Ahora bien, si sustituimos todo el tramo de curva que va desde A hasta $(0, y(T_0))$ por un tramo de caída vertical, obtenemos otra curva que tardará menos de T_0 unidades de tiempo en llegar al mismo punto (pues en vertical la aceleración es máxima, luego el tiempo es mínimo), luego la trayectoria dada no puede ser la braquistócrona que buscamos y podemos descartarla.

- El caso anterior no descarta la posibilidad de que la curva dada empiece con un tramo de caída en vertical, durante un intervalo $[0, T_0]$, de modo que $x(t) = 0$ si $0 \leq t \leq T_0 \leq T$ y $x(t) > 0$ si $t > T_0$.

En conclusión, podemos suponer que la trayectoria cumple $x(t) = 0$ para $0 \leq t \leq T_0$ (para cierto $0 \leq T_0 < T$) y $x(t) > 0$ para $t > T_0$ y también $y(t) \leq 0$ con la desigualdad estricta salvo en A y en posibles picos aislados.

Esto se traduce en que, para $T_0 < t \leq T$, podemos considerar las coordenadas cicloidales $(\rho(t), \tau(t))$ del punto $(x(t), -y(t))$, de modo que

$$\vec{r} = (x, y) = \left(\rho\tau - \rho^2 \operatorname{sen} \frac{\tau}{\rho}, -\rho^2(1 - \cos \frac{\tau}{\rho}) \right).$$

Más concretamente, $(\rho(t), \tau(t))$ es la composición de $(x(t), -y(t))$ con el cambio de variables inverso $f^{-1} :]0, +\infty[\times]0, +\infty[\rightarrow \mathbb{R}^2$ que asigna a cada punto (x, y) sus coordenadas cicloidales (ρ, τ) . Hemos probado que se trata de una aplicación de clase C^∞ , por lo que las funciones $\rho(t), \tau(t)$ son de clase C^1 en los puntos donde lo son $x(t), y(t)$ y además $y(t) < 0$. En la práctica sólo tenemos que incluir los instantes t en los que $y(t) = 0$ en el conjunto finito de instantes en los que \vec{r} no es derivable.

El teorema 5.18 garantiza que las funciones

$$\rho :]T_0, T] \rightarrow \mathbb{R}, \quad \tau : [T_0, T] \rightarrow \mathbb{R}$$

son continuas si en los posibles puntos donde $y(t) = 0$ las definimos mediante los límites dados por dicho teorema y hacemos $\tau(T_0) = \sqrt{-2y_0}$, donde y_0 es la altura hasta la que el cuerpo desciende en caída libre en el intervalo $[0, T_0]$. Ahora bien, puesto que en dicho intervalo la aceleración será g , dicha altura será concretamente $y_0 = -(1/2)gT_0^2$, luego $\tau(T_0) = \sqrt{g}T_0$.

Ahora calculamos la velocidad en coordenadas cicloidales (en los instantes $t > T_0$ donde la trayectoria es derivable):

$$\begin{aligned} x' &= \frac{\partial x}{\partial \rho} \rho' + \frac{\partial x}{\partial \tau} \tau' = \left(\tau - 2\rho \operatorname{sen} \frac{\tau}{\rho} + \tau \cos \frac{\tau}{\rho} \right) \rho' + \left(\rho - \rho \cos \frac{\tau}{\rho} \right) \tau', \\ y' &= \frac{\partial y}{\partial \rho} \rho' + \frac{\partial y}{\partial \tau} \tau' = -(2\rho - 2\rho \cos \frac{\tau}{\rho} - \tau \operatorname{sen} \frac{\tau}{\rho}) \rho' - \rho \operatorname{sen} \frac{\tau}{\rho} \tau', \\ x'^2 + y'^2 &= (8\rho^2(1 - \cos \frac{\tau}{\rho}) - 8\rho\tau \operatorname{sen} \frac{\tau}{\rho} + 2\tau^2(1 + \cos \frac{\tau}{\rho}))\rho'^2 + 2\rho^2(1 - \cos \frac{\tau}{\rho})\tau'^2 \\ &= 4(4\rho^2 \operatorname{sen}^2 \frac{\tau}{2\rho} - 4\rho\tau \operatorname{sen} \frac{\tau}{2\rho} \cos \frac{\tau}{2\rho} + \tau^2 \cos^2 \frac{\tau}{2\rho}) - 2y\tau'^2 \\ &= 4 \left(2\rho \operatorname{sen} \frac{\tau}{2\rho} - \tau \cos \frac{\tau}{2\rho} \right)^2 \rho'^2 - 2y\tau'^2. \end{aligned}$$

Así pues,

$$-2gy = v^2 = x'^2 + y'^2 = 16\rho^2 \left(\operatorname{sen} \frac{\tau}{2\rho} - \frac{\tau}{2\rho} \cos \frac{\tau}{2\rho} \right)^2 \rho'^2 - 2y\tau'^2. \quad (5.9)$$

Por consiguiente $-2y\tau'^2 \leq -2gy$ y, como $y < 0$, concluimos que $\tau' \leq \sqrt{g}$. Por lo tanto,⁹

$$\tau(T) - \sqrt{g}T_0 = \int_{T_0}^T \tau'(t) dt \leq \int_{T_0}^T \sqrt{g} dt = \sqrt{g}T - \sqrt{g}T_0,$$

luego el tiempo T necesario para llegar de A hasta B por cualquier trayectoria está acotado inferiormente por

$$T \geq \frac{\tau(B)}{\sqrt{g}}.$$

Más aún, se da la igualdad si y sólo si

$$\int_{T_0}^T \tau'(t) dt = \int_{T_0}^T \sqrt{g} dt,$$

o también, si y sólo si

$$\int_{T_0}^T (\sqrt{g} - \tau'(t)) dt = 0.$$

Pero sabemos que $\sqrt{g} - \tau'(t) \geq 0$, luego una primitiva F tiene que ser creciente, luego para que se cumpla $F(T) - F(T_0) = 0$ la primitiva tiene que ser constante, luego tiene que ser $\sqrt{g} - \tau'(t) = 0$.

En suma, una trayectoria que llegue desde A hasta B en el tiempo mínimo $\tau(B)/\sqrt{g}$ tiene que cumplir que $\tau' = \sqrt{g}$, pero por (5.9) esto equivale a que

$$\rho^2 \left(\operatorname{sen} \frac{\tau}{2\rho} - \frac{\tau}{2\rho} \cos \frac{\tau}{2\rho} \right)^2 \rho'^2 = 0,$$

pero $\rho > 0$ y, por la propiedad b) demostrada en la prueba del teorema 5.17, la expresión entre paréntesis es > 0 , luego tiene que ser $\rho' = 0$ o, lo que es lo mismo, ρ tiene que ser constante. Más aún, esto implica que $T_0 = 0$, porque en caso contrario tendríamos que

$$\lim_{t \rightarrow T_0^+} (x(t), y(t)) = (0, y_0).$$

con $y_0 < 0$, y el teorema 5.18 nos daría que $\lim_{t \rightarrow T_0^+} \rho(t) = +\infty$, lo cual es imposible si ρ es constante.

En resumen, el tiempo mínimo se consigue únicamente mediante las trayectorias de la forma

$$\rho = \rho_0, \quad \tau = \sqrt{g}t,$$

que son cicloides, y sólo hay una que pase por el punto B . Esto prueba que la braquistócrona es única y es la cicloide que tiene en A su punto más alto y pasa por B .

⁹Si hay valores de t donde $\tau'(t)$ no está definida, calculamos la integral descomponiéndola en suma de integrales en intervalos $[t_{i-1}, t_i]$ donde τ es continua y derivable en $]t_{i-1}, t_i[$.

Notemos que esto vale incluso si para llegar a B el objeto tiene que llegar al punto más bajo de la cicloide y volver a subir. Aun así, el tiempo empleado será menor que el que requiere cualquier otra trayectoria.

Observemos también que hemos obtenido la ecuación del movimiento de la partícula por la cicloide:

$$\vec{r}(t) = (\sqrt{Rg}t - R \operatorname{sen} \sqrt{\frac{g}{R}}t, -R(1 - \cos \sqrt{\frac{g}{R}}t)).$$

Ejemplo La cicloide de radio $R = 1$ corta a la recta $y = -x$ cuando

$$t - \operatorname{sen} t = 1 + \cos t,$$

lo cual sucede para $t = 0$ y $t = 2.412$, y el punto de corte es $(1.745, -1.745)$. Tomando $R = 100/1.745 = 57.292$ conseguimos una cicloide que pasa por el punto $B = (1, -1)$ cuando $t = 2.412$. Las coordenadas cicloidales de B son

$$\rho = \sqrt{R} = 7.569, \quad \tau = t\rho = 18.257$$

Por lo tanto, el tiempo que tarda un objeto en caer desde $A = (0, 0)$ hasta $B = (1, -1)$ a lo largo de la braquistócrona es

$$T = \frac{\tau(B)}{\sqrt{g}} = 5.832 \text{ s.}$$

La tabla siguiente contiene los tiempos que hemos calculado en ejemplos precedentes para curvas distintas:

	Recta	Dos rectas	Circunferencia	Braquistócrona
T	6.39	5.99	5.92	5.83

■

Capítulo VI

Integrales y funciones elípticas

Hemos probado que toda función continua tiene primitiva, pero hay muchas funciones continuas sencillas, como e^{x^2} , cuya primitiva no puede expresarse en términos de funciones “conocidas”. En lugar de interpretar esto como que una primitiva de tales funciones es una función “misteriosa”, es más razonable pensar que la integración permite definir nuevas funciones a partir de otras ya conocidas. Un ejemplo que ya hemos visto de esta situación es la función factorial, que puede definirse a través de una integral impropia. Aquí vamos a estudiar algunos ejemplos de las que Legendre llamó *integrales elípticas*, pues una de ellas es la que permite calcular la longitud de un arco de elipse.

6.1 Las integrales de primera y segunda especie

La elipse de ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

donde $0 < b \leq a$, admite la expresión paramétrica

$$r(t) = (a \operatorname{sen} \phi, b \operatorname{cos} \phi), \quad 0 \leq \phi \leq 2\pi.$$

De aquí obtenemos que su longitud es

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi} d\phi = 4 \int_0^{\pi/2} \sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi} d\phi \\ &= 4 \int_0^{\pi/2} \sqrt{a^2 - (a^2 - b^2) \operatorname{sen}^2 \phi} d\phi = 4a \int_0^{\pi/2} \sqrt{1 - \epsilon^2 \operatorname{sen}^2 \phi} d\phi, \end{aligned}$$

donde $\epsilon = \sqrt{(a^2 - b^2)/a^2}$ es la excentricidad, que cumple $0 \leq \epsilon < 1$.

Para $\epsilon = 0$ obtenemos $L = 2\pi a$, que es la longitud de la circunferencia de radio a , pero para elipses propiamente dichas la integral no puede expresarse en términos de funciones trigonométricas, exponenciales, etc.

En general, para $|m| < 1$, el cambio de variable $t = \text{sen } \phi$ nos da que

$$\int_0^\theta \sqrt{1 - m \text{sen}^2 \phi} d\phi = \int_0^x \frac{\sqrt{1 - mt^2}}{\sqrt{1 - t^2}} dt,$$

donde $x = \text{sen } \theta$, y la integral

$$E(x; m) = \int_0^x \frac{\sqrt{1 - mt^2}}{\sqrt{1 - t^2}} dt = \int_0^x \frac{\sqrt{(1 - mt^2)(1 - t^2)}}{1 - t^2} dt$$

es la que Legendre definió como *integral elíptica de segunda especie*. También se define la versión trigonométrica

$$E(\theta | m) = \int_0^\theta \sqrt{1 - m \text{sen}^2 \phi} d\phi,$$

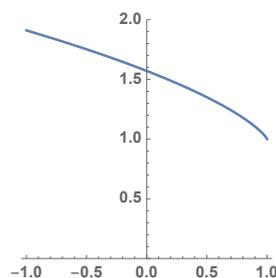
de modo que¹ $E(\theta | m) = E(\text{sen } \theta; m)$.

La *integral completa de segunda especie* se define como

$$E(m) = \int_0^1 \frac{\sqrt{1 - mt^2}}{\sqrt{1 - t^2}} dt = \int_0^{\pi/2} \sqrt{1 - m \text{sen}^2 \phi} d\phi.$$

En estos términos, hemos demostrado que la longitud de una elipse de semieje mayor a y excentricidad ϵ es $L = 4a E(\epsilon^2)$.

La figura muestra la gráfica de $E(m)$. Es claro que $E(0) = \pi/2$ y que $E(1) = 1$, pues se trata de la integral del coseno. Para valores del parámetro $|m| < 1$ puede calcularse mediante una serie de potencias:



Teorema 6.1 Si $|m| < 1$, se cumple que

$$E(m) = \frac{\pi}{2} \sum_{n=0}^{\infty} \left(\prod_{j=1}^n \frac{2j-1}{2j} \right)^2 \frac{m^n}{1-2n} =$$

$$\frac{\pi}{2} \left(1 - \left(\frac{1}{2} \right)^2 \frac{m}{1} - \left(\frac{1 \cdot 3}{2 \cdot 4} \right)^2 \frac{m^2}{3} - \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \right)^2 \frac{m^3}{5} - \dots \right).$$

¹Es fácil encontrar sutiles cambios de notación en trabajos distintos sobre integrales elípticas, y a menudo las notaciones son ambiguas. Nosotros, por ejemplo, distinguiremos mediante el uso del punto y coma o de la barra si el primer argumento es x o θ . Otro punto en el que el lector debe prestar atención si consulta otras fuentes es que algunas llaman $m = k^2$ y definen $E(\theta | k) = \int_0^\theta \sqrt{1 - k^2 \text{sen}^2 \phi} d\phi$, de forma que $E(\theta | k)$ es lo que nosotros llamamos $E(\theta | k^2)$. El argumento m se llama *parámetro*, mientras que k se llama *módulo* o *excentricidad*.

DEMOSTRACIÓN: Usamos el teorema 1.23 para desarrollar la raíz cuadrada:

$$\begin{aligned} E(m) &= \int_0^{\pi/2} \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{3/2-j}{j} (-1)^n m^n \operatorname{sen}^{2n} \phi \, d\phi \\ &= \int_0^{\pi/2} \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{2j-3}{2j} m^n \operatorname{sen}^{2n} \phi \, d\phi. \end{aligned}$$

Ahora observamos que todos los factores del producto son positivos excepto el correspondiente a $j = 1$, luego todos los productos son negativos excepto el correspondiente a $n = 0$, que vale 1. Por lo tanto, la sucesión de sumas parciales

$$-\sum_{n=1}^N \prod_{j=1}^n \frac{2j-3}{2j} m^n \operatorname{sen}^{2n} \phi$$

es creciente, y podemos aplicar el teorema 2.8 para concluir que

$$-\sum_{n=1}^{\infty} \prod_{j=1}^n \frac{2j-3}{2j} m^n \int_0^{\pi/2} \operatorname{sen}^{2n} \phi \, d\phi = -\int_0^{\pi/2} \sum_{n=1}^{\infty} \prod_{j=1}^n \frac{2j-3}{2j} m^n \operatorname{sen}^{2n} \phi \, d\phi,$$

luego, quitando el signo negativo e incorporando el primer sumando:

$$E(m) = \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{2j-3}{2j} \int_0^{\pi/2} \operatorname{sen}^{2n} \phi \, d\phi \, m^n.$$

Ahora usamos la relación (C.9), según la cual

$$\int_0^{\pi/2} \operatorname{sen}^{2n} \phi \, d\phi = \frac{2n-1}{2n} \int_0^{\pi/2} \operatorname{sen}^{2(n-1)} \phi \, d\phi,$$

luego, aplicándola n veces:

$$\int_0^{\pi/2} \operatorname{sen}^{2n} \phi \, d\phi = \prod_{j=1}^n \frac{2j-1}{2j} \int_0^{\pi/2} d\phi = \frac{\pi}{2} \prod_{j=1}^n \frac{2j-1}{2j}.$$

Por lo tanto:

$$E(m) = \frac{\pi}{2} \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{2j-3}{2j} \prod_{j=1}^n \frac{2j-1}{2j} m^n.$$

El sumando correspondiente a $n = 0$ vale 1, mientras que, en los siguientes, el primer producto tiene los mismos denominadores que el segundo, y a los numeradores les sobra el -1 inicial y les falta el $2n - 1$ final, luego

$$\begin{aligned} E(m) &= \frac{\pi}{2} + \frac{\pi}{2} \sum_{n=1}^{\infty} \frac{-1}{2n-1} \left(\prod_{j=1}^n \frac{2j-1}{2j} \right)^2 m^n = \\ &= \frac{\pi}{2} + \frac{\pi}{2} \sum_{n=1}^{\infty} \left(\prod_{j=1}^n \frac{2j-1}{2j} \right)^2 \frac{m^n}{1-2n} \end{aligned}$$

que es la expresión del enunciado. ■

Un enunciado equivalente del teorema anterior es:

Teorema 6.2 La longitud de una elipse de semieje mayor a y excentricidad ϵ es

$$L = 2\pi a \sum_{n=0}^{\infty} \left(\prod_{j=1}^n \frac{2j-1}{2j} \right)^2 \frac{\epsilon^{2n}}{1-2n} =$$

$$2\pi a \left(1 - \left(\frac{1}{2}\right)^2 \frac{\epsilon^2}{1} - \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 \frac{\epsilon^4}{3} - \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\right)^2 \frac{\epsilon^6}{5} - \dots \right).$$

Ejemplo La convergencia no es muy rápida. Por ejemplo, la tabla siguiente muestra las sumas parciales de la serie de $E(3/4)$:

0	1.57079632679	11	1.21121665937
1	1.27627201552	12	1.21116045748
2	1.23485453425	13	1.21112460384
3	1.22191157135	14	1.21110145217
4	1.21660293422	15	1.21108634571
5	1.21409460318	16	1.21107639890
6	1.21280124499	17	1.21106979709
7	1.21209352730	18	1.21106538443
8	1.21168921593	19	1.21106241642
9	1.21145055991	20	1.21106040884
10	1.21130602386	21	1.21105904399
		∞	1.21105602757...

Se necesitan 9 términos para conseguir 3 decimales exactos y 16 para conseguir 4. En particular, vemos que la longitud de una elipse de semiejes $a = 2$ y $b = 1$ (luego excentricidad $\epsilon = \sqrt{3}/2$) es $L = 4 \cdot 2E(3/4) = 9.6884\dots$ ■

La longitud del seno Otro contexto en el que aparece la función E es al calcular la longitud de un arco de la función seno:

$$L = \int_0^{\pi} \sqrt{1 + \cos^2 x} dx = 2 \int_0^{\pi/2} \sqrt{2 - \operatorname{sen}^2 x} dx$$

$$= 2\sqrt{2} \int_0^{\pi/2} \sqrt{1 - (1/2)\operatorname{sen}^2 x} dx = 2\sqrt{2}E(1/2) = 3.8201977890\dots$$

donde hemos usado que $E(1/2) = 1.35064388104767550\dots$ ■

Legendre definió la *integral elíptica de primera especie* como

$$F(x; m) = \int_0^x \frac{1}{\sqrt{(1 - mt^2)(1 - t^2)}} dt,$$

o bien, en forma trigonométrica,

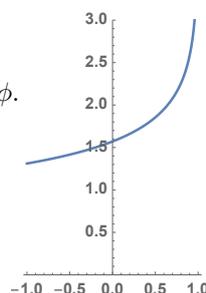
$$F(\theta | m) = \int_0^{\theta} \frac{1}{1 - m \operatorname{sen}^2 \phi} d\phi,$$

de modo que el cambio de variable $t = \operatorname{sen} \phi$ muestra que $F(\theta | m) = F(\operatorname{sen} \theta; m)$.

La *integral completa de primera especie* se define como

$$K(m) = \int_0^1 \frac{1}{\sqrt{(1-mt^2)(1-t^2)}} dt = \int_0^{\pi/2} \frac{1}{\sqrt{1-m \operatorname{sen}^2 \phi}} d\phi.$$

Notemos que la primera de estas dos integrales es impropia en 1, pero la segunda no lo es, y la igualdad entre las dos expresiones de la integral incompleta para todo $0 \leq \theta < \pi/2$ implica la convergencia de la primera integral.



La figura muestra la gráfica de $K(m)$. Es inmediato que $K(0) = \pi/2$, mientras que $\lim_{m \rightarrow 1} K(m) = +\infty$.

Una ligera variante de la prueba del teorema 6.1 nos da su desarrollo en serie de potencias:

Teorema 6.3 Si $|m| < 1$, se cumple que

$$K(m) = \frac{\pi}{2} \sum_{n=0}^{\infty} \left(\prod_{j=1}^n \frac{2j-1}{2j} \right)^2 m^n = \frac{\pi}{2} \left(1 + \left(\frac{1}{2}\right)^2 m + \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 m^2 + \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\right)^2 m^3 + \dots \right).$$

DEMOSTRACIÓN: Usamos el teorema 1.23 para desarrollar la raíz cuadrada:

$$\begin{aligned} K(m) &= \int_0^{\pi/2} \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{1/2-j}{j} (-1)^n m^n \operatorname{sen}^{2n} \phi d\phi \\ &= \int_0^{\pi/2} \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{2j-1}{2j} m^n \operatorname{sen}^{2n} \phi d\phi. \end{aligned}$$

Como todos los sumandos son positivos, el teorema 2.8 nos permite intercambiar la integral y la suma, y la integral del seno se calcula como en 6.1:

$$\begin{aligned} K(m) &= \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{2j-1}{2j} \int_0^{\pi/2} \operatorname{sen}^{2n} \phi d\phi m^n = \\ &= \frac{\pi}{2} \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{2j-1}{2j} \prod_{j=1}^n \frac{2j-1}{2j} m^n, \end{aligned}$$

que es la fórmula del enunciado. ■

Por ejemplo, $K(1/2) = 1.854074677301371 \dots$ y para conseguir 10 decimales exactos hay que sumar 36 términos de la serie.

Hay varias integrales interesantes que pueden reducirse a $K(1/2)$:

Teorema 6.4

$$\int_0^1 \frac{1}{\sqrt{1+x^4}} dx = \frac{1}{2}K(1/2), \quad \int_0^1 \frac{1}{\sqrt{1-x^4}} dx = K(-1) = \frac{\sqrt{2}}{2}K(1/2),$$

$$\int_0^{\pi/2} \frac{1}{\sqrt{\sin x}} dx = \int_0^{\pi/2} \frac{1}{\sqrt{\cos x}} dx = \sqrt{2}K(1/2).$$

DEMOSTRACIÓN: Tenemos que

$$K(1/2) = \int_0^{\pi/2} \frac{1}{\sqrt{1-(1/2)\sin^2 \phi}} d\phi = \sqrt{2} \int_0^{\pi/2} \frac{1}{\sqrt{2-\sin^2 \phi}} d\phi.$$

Si aplicamos el cambio de variable

$$x = \tan(\phi/2), \quad \phi = 2 \arctan x, \quad d\phi = \frac{2}{1+x^2} dx,$$

$$2 - \sin^2 \phi = 2 - \frac{4x^2}{(1+x^2)^2} = 2 \frac{1+x^4}{(1+x^2)^2},$$

obtenemos

$$K(1/2) = \sqrt{2} \int_0^1 \frac{1}{\sqrt{2} \frac{\sqrt{1+x^4}}{1+x^2}} \frac{2}{1+x^2} dt = 2 \int_0^1 \frac{1}{\sqrt{1+x^4}} dt.$$

Por otro lado, el cambio

$$x = \cos \phi, \quad dx = -\sqrt{1-x^2} d\phi$$

nos da

$$K(1/2) = \sqrt{2} \int_0^1 \frac{1}{\sqrt{1+x^2}} \frac{dx}{\sqrt{1-x^2}} = \sqrt{2} \int_0^1 \frac{1}{\sqrt{1-x^4}} dx.$$

La igualdad

$$\int_0^1 \frac{1}{\sqrt{1-x^4}} dx = K(-1)$$

es inmediata por la definición de $K(m)$.

La tercera integral del enunciado es impropia en 0. Para probar su convergencia fijamos $\epsilon > 0$ y observamos que

$$\int_{\epsilon}^{\pi/2} \frac{1}{\sqrt{\sin x}} dx = 2 \int_{\epsilon/2}^{\pi/4} \frac{1}{\sqrt{\sin 2x}} dx = \sqrt{2} \int_{\epsilon/2}^{\pi/4} \frac{1}{\sqrt{\sin x \cos x}} dx,$$

y ahora hacemos el cambio de variable

$$\tan x = t^2, \quad x = \arctan t^2, \quad dx = \frac{2t}{1+t^4} dt,$$

$$\sin x \cos x = \frac{\sin x}{\cos x} \cos^2 x = \tan x \frac{1}{1/\cos^2 x} = \frac{\tan x}{1+\tan^2 x} = \frac{t^2}{1+t^4}.$$

Así la integral se convierte en

$$\sqrt{2} \int_{\sqrt{\tan(\epsilon/2)}}^1 \frac{\sqrt{1+t^4}}{t} \frac{1}{1+t^4} 2t dt = 2\sqrt{2} \int_{\sqrt{\tan(\epsilon/2)}}^1 \frac{1}{\sqrt{1+t^4}} dt,$$

y ahora basta tomar límites cuando ϵ tiende a 0 para concluir que

$$\int_0^{\pi/2} \frac{1}{\sqrt{\sin x}} dx = 2\sqrt{2} \int_0^1 \frac{1}{\sqrt{1+t^4}} dt = \sqrt{2}K(1/2).$$

El cambio de variable $x = \pi/2 - x'$ nos da que

$$\int_0^{\pi/2-\epsilon} \frac{1}{\sqrt{\cos x}} dx = \int_{\epsilon}^{\pi/2} \frac{1}{\sqrt{\sin x}} dx,$$

y al tomar límites obtenemos la igualdad entre las dos últimas integrales del enunciado. ■

En particular, según el ejemplo de la página 334, el tiempo que tarda un objeto en caer por un cuarto de circunferencia de radio R es

$$T = \sqrt{\frac{R}{g}} K(1/2) \approx 0.189191\sqrt{R}.$$

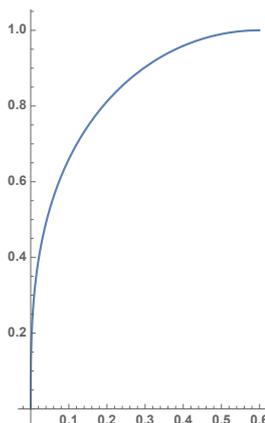
Por otro lado, la relación (3.6) conecta la función factorial con la función K :

$$\Gamma^2(1/4) = \frac{\sqrt{\pi}}{4} K(1/2).$$

Como hemos visto en la introducción, Jakob Bernoulli se encontró estas integrales en el estudio de la curva elástica rectangular, lo que a su vez le llevó a descubrir la lemniscata:

La curva elástica rectangular Analizamos ahora el problema de la elástica rectangular, planteado por Jakob Bernoulli en 1691, es decir, el problema de determinar la forma que adquiere una vara elástica cuando se fija un extremo al suelo y se fija un peso en el otro extremo de la magnitud adecuada para que la curva 90° , tal y como muestra la figura.

Bernoulli la encontró partiendo de que la curvatura en cada punto debe ser proporcional a la altura. Éste es un principio que probablemente Bernoulli constató empíricamente, cuya naturaleza es similar a la ley de Hooke. Justificarlo requeriría entrar en consideraciones de la mecánica de los sólidos que exceden el nivel de este libro, pero, aun aceptándolo como hizo Bernoulli, encontrar la curva requería cierto nivel matemático, sobre todo porque el concepto de “curvatura” no estaba muy desarrollado entonces.



Vamos a encontrar una parametrización de la forma $(x(y), y)$, de modo que un extremo de la vara sea $(0, 0)$. Usando la última de las fórmulas (4.1) obtenemos fácilmente que la curvatura es

$$\kappa = \frac{x''}{(x'^2 + 1)^{3/2}},$$

por lo que la condición de Bernoulli es:

$$\frac{x''}{(x'^2 + 1)^{3/2}} = \frac{2}{h^2}y,$$

donde por conveniencia expresamos la constante de proporcionalidad como $2/h^2$. Esta ecuación diferencial de segundo orden se convierte en una de primer orden con el cambio $u(y) = x'(y)$:

$$\frac{du}{(1 + u^2)^{3/2}} = \frac{2y}{h^2} dy,$$

que al integrar, teniendo en cuenta (C.18), se convierte en

$$\frac{u}{\sqrt{1 + u^2}} = (y/h)^2 + c$$

o, equivalentemente,

$$\frac{x'}{\sqrt{1 + x'^2}} = (y/h)^2 + c.$$

Pero el hecho de que la vara tenga un extremo perpendicular al eje x se traduce en que $x'(0) = 0$, luego $c = 0$ y, al despejar x' , resulta que

$$\frac{dx}{dy} = \frac{(y/h)^2}{\sqrt{1 - (y/h)^4}}.$$

Por lo tanto, la curva elástica rectangular es la gráfica de la función

$$x(y) = \int_0^y \frac{(t/h)^2}{\sqrt{1 - (t/h)^4}} dt = h \int_0^{y/h} \frac{t^2}{\sqrt{1 - t^4}} dt.$$

En particular, el desplazamiento horizontal de una elástica rectangular de altura h es

$$D = h \int_0^1 \frac{t^2}{\sqrt{1 - t^4}} dt.$$

Ahora se ve por qué hemos expresado la constante de proporcionalidad como lo hemos hecho, y es que así h es la altura de la varilla doblada (el valor de y que hace la derivada x' infinita, es decir, la que hace que la tangente sea horizontal), y está determinada a su vez por la longitud de la varilla, que resulta ser

$$\begin{aligned} L &= \int_0^h \sqrt{x'^2 + 1} dy = \int_0^h \sqrt{\frac{(y/h)^4}{1 - (y/h)^4} + 1} dy = \int_0^h \frac{1}{\sqrt{1 - (y/h)^4}} dy \\ &= h \int_0^1 \frac{1}{\sqrt{1 - t^4}} dt = h \frac{\sqrt{2}}{2} K(1/2). \end{aligned}$$

En realidad es al revés: si la varilla recta tiene longitud L , al doblarse 90° su altura pasa a ser

$$h = \frac{\sqrt{2}L}{K(1/2)} \approx 0.76276L$$

y, según la relación (3.7), que equivale a que

$$LD = h^2 \int_0^1 \frac{1}{\sqrt{1-t^4}} dt \int_0^1 \frac{x^2}{\sqrt{1-t^4}} dt = h^2 \frac{\pi}{4},$$

el desplazamiento es

$$D = \frac{\pi L}{2K^2(1/2)} \approx 0.456947L.$$

Tal y como hemos explicado en la introducción, Jakob Bernoulli, consciente de que no podía calcular de forma exacta la integral que determina la elástica, trató de interpretarla geoméricamente representándola como la longitud de una curva sencilla, y así encontró la lemniscata que lleva su nombre. En efecto, en la sección B.5 vemos que su longitud respecto de su parametrización en función de ρ viene dada por la misma función $1/\sqrt{1-t^4}$, pero, al contrario que la elástica rectangular, la lemniscata admite una ecuación algebraica y una parametrización algebraica. ■

Ahora calculamos las derivadas de las integrales elípticas:

Teorema 6.5 Si $0 < m < 1$, se cumple que

$$\frac{dK}{dm} = \frac{E - (1-m)K}{2m(1-m)}, \quad \frac{dE}{dm} = \frac{E - K}{2m}.$$

DEMOSTRACIÓN: La derivada de la derecha es la más fácil de calcular. Basta aplicar el teorema 3.11, aunque sin necesidad de acotar la derivada del integrando, pues la integral no es impropia:

$$\frac{dE}{dm} = \frac{d}{dm} \int_0^{\pi/2} \sqrt{1 - m \operatorname{sen}^2 \phi} d\phi = -\frac{1}{2} \int_0^{\pi/2} \frac{\operatorname{sen}^2 \phi}{\sqrt{1 - m \operatorname{sen}^2 \phi}} d\phi$$

y, por otra parte,

$$\begin{aligned} E - K &= \int_0^{\pi/2} \left(\sqrt{1 - m \operatorname{sen}^2 \phi} - \frac{1}{\sqrt{1 - m \operatorname{sen}^2 \phi}} \right) d\phi \\ &= \int_0^{\pi/2} \frac{-m \operatorname{sen}^2 \phi}{\sqrt{1 - m \operatorname{sen}^2 \phi}} d\phi, \end{aligned}$$

luego se cumple la igualdad del enunciado. Una vez probada ésta, la primera es equivalente a

$$\frac{d(E - K)}{dm} = -\frac{E}{2(1-m)}. \quad (6.1)$$

Todo se reduce a una comprobación laboriosa:

$$\frac{d(E - K)}{dm} = \int_0^{\pi/2} \left(\frac{-\operatorname{sen}^2 \phi}{\sqrt{1 - m \operatorname{sen}^2 \phi}} - \frac{m \operatorname{sen}^4 \phi}{2(1 - m \operatorname{sen}^2 \phi)^{3/2}} \right) d\phi.$$

Operamos el integrando:

$$\begin{aligned} \frac{-2 \operatorname{sen}^2 \phi (1 - m \operatorname{sen}^2 \phi) - m \operatorname{sen}^4 \phi}{2(1 - m \operatorname{sen}^2 \phi)^{3/2}} &= \frac{-2 \operatorname{sen}^2 \phi + m \operatorname{sen}^4 \phi}{2(1 - m \operatorname{sen}^2 \phi)^{3/2}} = \\ \frac{-2m \operatorname{sen}^2 \phi + m^2 \operatorname{sen}^4 \phi}{2m(1 - m \operatorname{sen}^2 \phi)^{3/2}} &= \frac{(1 - m \operatorname{sen}^2 \phi)^2 - 1}{2m(1 - m \operatorname{sen}^2 \phi)^{3/2}} = \\ \frac{1}{2m} \left(\sqrt{1 - m \operatorname{sen}^2 \phi} - \frac{1}{(1 - m \operatorname{sen}^2 \phi)^{3/2}} \right). \end{aligned}$$

Así pues:

$$\frac{d(E - K)}{dm} = \frac{E}{2m} - \frac{1}{2m} \int_0^{\pi/2} (1 - m \operatorname{sen}^2 \phi)^{-3/2} d\phi.$$

Por otra parte

$$\begin{aligned} \frac{d}{d\phi} \frac{\operatorname{sen} \phi \cos \phi}{(1 - m \operatorname{sen}^2 \phi)^{1/2}} &= \frac{1 - 2 \operatorname{sen}^2 \phi + m \operatorname{sen}^4 \phi}{(1 - m \operatorname{sen}^2 \phi)^{3/2}} = \\ \frac{m - 2m \operatorname{sen}^2 \phi + m^2 \operatorname{sen}^4 \phi}{m(1 - m \operatorname{sen}^2 \phi)^{3/2}} &= \frac{(1 - m \operatorname{sen}^2 \phi)^2 - (1 - m)}{m(1 - m \operatorname{sen}^2 \phi)^{3/2}} = \\ \frac{1}{m} \sqrt{1 - m \operatorname{sen}^2 \phi} - \frac{1 - m}{m} (1 - m \operatorname{sen}^2 \phi)^{-3/2}. \end{aligned}$$

Al integrar queda

$$0 = \frac{E}{m} - \frac{1 - m}{m} \int_0^{\pi/2} (1 - m \operatorname{sen}^2 \phi)^{-3/2} d\phi,$$

luego

$$\int_0^{\pi/2} (1 - m \operatorname{sen}^2 \phi)^{-3/2} d\phi = \frac{E}{1 - m}$$

y concluimos que

$$\frac{d(E - K)}{dm} = \frac{E}{2m} - \frac{1}{2m} \frac{E}{1 - m} = \frac{E}{2m} \left(1 - \frac{1}{1 - m} \right) = -\frac{E}{2(1 - m)}.$$

■

Como aplicación probamos una de las muchas relaciones que Legendre encontró entre las integrales elípticas:

Teorema 6.6 (Relación de Legendre) Si $0 < m < 1$, entonces

$$K(m)E(1 - m) + K(1 - m)E(m) - K(m)K(1 - m) = \frac{\pi}{2}.$$

DEMOSTRACIÓN: Si llamamos X al miembro izquierdo de la igualdad del enunciado, entonces

$$X(m) = E(m)E(1-m) - (E(m) - K(m))(E(1-m) - K(1-m)).$$

Por el teorema anterior, considerando también (6.1), tenemos que

$$\begin{aligned} \frac{dX}{dm} &= \frac{E(m) - K(m)}{2m} E(1-m) - E(m) \frac{E(1-m) - K(1-m)}{2(1-m)} \\ &+ \frac{E(m)}{2(1-m)} (E(1-m) - K(1-m)) + (E(m) - K(m)) \frac{E(1-m)}{2m} = 0. \end{aligned}$$

Así pues, X es constante. Ahora basta probar que $\lim_{m \rightarrow 0} X(m) = \pi/2$.

Recordemos que $E(0) = K(0) = \pi/2$, $E(1) = 1$, y el único problema es que $\lim_{m \rightarrow 1} K(m) = +\infty$. Así pues, hay que probar que

$$\lim_{m \rightarrow 0} K(1-m)(E(m) - K(m)) = 0.$$

En efecto, si $m \leq \pi/2$, se cumple que

$$\begin{aligned} |E(m) - K(m)| &= \left| \int_0^{\pi/2} \frac{-m \operatorname{sen}^2 \phi}{\sqrt{1 - m \operatorname{sen}^2 \phi}} d\phi \right| \leq m \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - m \operatorname{sen}^2 \phi}} \leq \\ &m \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - 0.5 \operatorname{sen}^2 \phi}} = Cm \end{aligned}$$

y

$$|K(1-m)| = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - (1-m) \operatorname{sen}^2 \phi}} \leq \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - (1-m)}} = \frac{\pi}{2\sqrt{m}}.$$

Por lo tanto,

$$|K(1-m)(E(m) - K(m))| \leq \frac{C\pi\sqrt{m}}{2},$$

lo que prueba que el límite vale 0. ■

Ejemplo Para $m = 1/2$ la relación de Legendre se reduce a

$$(2E(1/2) - K(1/2))K(1/2) = \frac{\pi}{2}.$$

Según el teorema 6.4,

$$K(1/2) = \sqrt{2} \int_0^1 \frac{dx}{\sqrt{1-x^4}},$$

mientras que el cambio de variable $x = \cos \phi$ nos da que

$$E(1/2) = \frac{\sqrt{2}}{2} \int_0^{\pi/2} \frac{d\phi}{\sqrt{2 - \sin^2 \phi}} = \frac{\sqrt{2}}{2} \int_0^1 \frac{1 + x^2}{\sqrt{1 - x^4}} dx,$$

luego

$$2E(1/2) - K(1/2) = \sqrt{2} \int_0^1 \frac{x^2}{\sqrt{1 - x^4}} dx$$

y la relación de Legendre para $m = 1/2$ expresa que

$$\int_0^1 \frac{1}{\sqrt{1 - x^4}} dx \int_0^1 \frac{x^2}{\sqrt{1 - x^4}} dx = \frac{\pi}{4},$$

que es la misma relación (3.7) que ya habíamos demostrado. ■

6.2 La lemniscata de Bernoulli

Vamos a analizar ahora los resultados que obtuvo Fagnano sobre la lemniscata de Bernoulli. Para ello recordemos de la sección B.5 que una parametrización del primer cuadrante de la lemniscata (tomando por simplicidad $a = 1$) viene dada por

$$r(\rho) = \left(\sqrt{\frac{1}{2}\rho^2(1 + \rho^2)}, \sqrt{\frac{1}{2}\rho^2(1 - \rho^2)} \right), \quad 0 \leq \rho \leq 1,$$

de modo que la longitud del arco que va desde $(0, 0)$ hasta $r(u)$ es

$$L(u) = \int_0^u \frac{d\rho}{\sqrt{1 - \rho^4}}.$$

Vamos a aplicar el cambio de variable

$$\rho^2 = \frac{2t^2}{1 + t^4}, \quad t^2 = \frac{1 - \sqrt{1 - \rho^4}}{\rho^2}.$$

(Notemos que el signo correcto ante la raíz cuadrada es el negativo para que tanto $\rho(t)$ como $t(\rho)$ sean funciones $[0, 1] \rightarrow [0, 1]$.)

Un cálculo rutinario muestra que

$$\rho d\rho = \frac{2t(1 - t^4)}{(1 + t^4)^2} dt, \quad d\rho = \frac{\sqrt{2}(1 - t^4)}{(1 + t^4)^{3/2}} dt,$$

con lo que, llamando $w = t(u)$, de modo que

$$w^2 = \frac{1 - \sqrt{1 - u^4}}{u^2},$$

tenemos que

$$\int_0^u \frac{d\rho}{\sqrt{1 - \rho^4}} = \sqrt{2} \int_0^w \frac{dt}{\sqrt{1 + t^4}}.$$

A la integral obtenida le aplicamos el cambio $t : [0, \sqrt{\sqrt{2}-1}] \rightarrow [0, 1]$ dado por

$$t^2 = \frac{2\rho^2}{1-\rho^4}, \quad \rho^2 = \frac{-1 + \sqrt{1+t^4}}{t^2}.$$

Ahora

$$t dt = \frac{2\rho(1+\rho^4)}{(1-\rho^4)^2} d\rho, \quad dt = \frac{\sqrt{2}(1+\rho^4)}{(1-\rho^4)^{3/2}} d\rho,$$

con lo que, llamando $v = \rho(w)$,

$$L(u) = \int_0^u \frac{d\rho}{\sqrt{1-\rho^4}} = 2 \int_0^v \frac{d\rho}{\sqrt{1-\rho^4}} = 2L(v).$$

Así pues, v es el valor de ρ que determina un arco de lemniscata de longitud igual a la mitad de la del arco determinado por u . Los cálculos nos dan una expresión explícita para la función $v : [0, 1] \rightarrow [0, \sqrt{\sqrt{2}-1}]$:

$$\begin{aligned} v^2 &= \frac{-1 + \sqrt{1+w^4}}{w^2} = \frac{-1 + \sqrt{1+w^4}}{w^2} = \frac{-1 + \sqrt{1 + ((1 - \sqrt{1-u^4})/u^2)^2}}{(1 - \sqrt{1-u^4})/u^2} \\ &= \frac{-u^2 + \sqrt{2 - 2\sqrt{1-u^4}}}{1 - \sqrt{1-u^4}}. \end{aligned}$$

El teorema siguiente resume lo que hemos demostrado:

Teorema 6.7 Consideremos la parametrización de un cuarto de lemniscata

$$r(\rho) = \left(\sqrt{\frac{1}{2}\rho^2(1+\rho^2)}, \sqrt{\frac{1}{2}\rho^2(1-\rho^2)} \right), \quad 0 \leq \rho \leq 1,$$

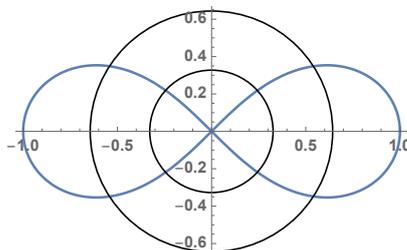
y sea $v : [0, 1] \rightarrow [0, \sqrt{\sqrt{2}-1}]$ dada por

$$v(u) = \sqrt{\frac{-u^2 + \sqrt{2 - 2\sqrt{1-u^4}}}{1 - \sqrt{1-u^4}}}.$$

Entonces las longitudes de los arcos determinados por u y $v = v(u)$ cumplen

$$L(u) = \int_0^u \frac{d\rho}{\sqrt{1-\rho^4}} = 2 \int_0^v \frac{d\rho}{\sqrt{1-\rho^4}} = 2L(v).$$

La expresión de $v(u)$ puede ser un poco farragosa, pero lo notable es que es algebraica. Más aún, los números $v(1) = \sqrt{\sqrt{2}-1}$, $v(v(1))$, etc. se pueden construir con regla y compás, lo que nos permite determinar segmentos de longitud igual a la mitad, la cuarta parte, etc. de un cuarto de lemniscata:



En la figura, el círculo mayor tiene radio $v(1)$, con lo que, junto con los ejes, divide la lemniscata en ocho arcos de la misma longitud. El círculo menor tiene radio $v(v(1))$, con lo que divide cuatro de estos arcos en ocho de la mitad de longitud. Para tener 16 arcos iguales, nos falta dividir los otros cuatro. Para ello Fagnano contaba con otro resultado:

Teorema 6.8 *Para la parametrización de la lemniscata considerada en el teorema anterior, consideramos ahora $w : [0, 1] \rightarrow [0, 1]$ dada por*

$$w(u) = \sqrt{\frac{1-u^2}{1+u^2}}.$$

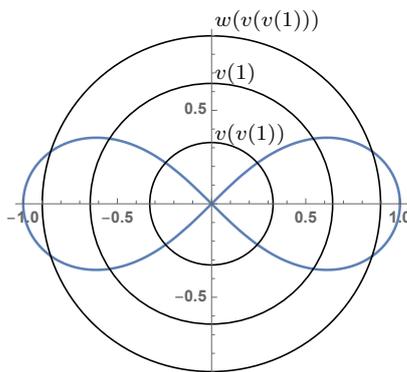
Entonces, para $w = w(u)$, se cumple que

$$\int_0^u \frac{d\rho}{\sqrt{1-\rho^4}} = \int_w^1 \frac{d\rho}{\sqrt{1-\rho^4}}.$$

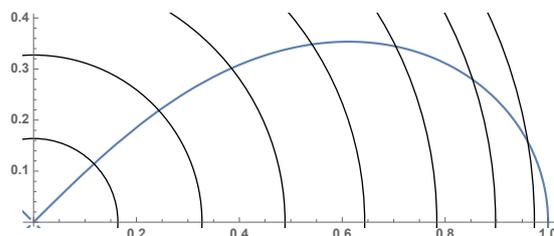
DEMOSTRACIÓN: La prueba es completamente análoga a la del teorema anterior, sin más que observar que el cambio de variable inverso es

$$u^2 = \frac{1-v^2}{1+v^2}. \quad \blacksquare$$

Esto significa que el arco comprendido entre $\rho = 0$ y $\rho = u$ tiene la misma longitud que el comprendido entre $\rho = w$ y $\rho = 1$. Así, trazando la circunferencia de radio $w(v(v(1)))$ obtenemos los arcos que nos faltaban para dividir la lemniscata en 16 partes iguales:



Si queremos dividirla en 32 partes iguales basta aplicar el teorema 6.7 a los segmentos impares, es decir, calcular $v(v(v(1)))$ y $v(w(v(v(1))))$, con lo que bisebamos los 8 arcos más cercanos a $(0, 0)$ y luego aplicar el teorema 6.8 a los radios obtenidos, con lo que dividimos los arcos restantes:



De este modo Fagnano podía medir lemniscatas, como afirmaba (tomando como unidad de medida la longitud completa), pues, dado cualquier arco en una lemniscata, podemos dividirla en 2^n partes de la misma longitud y ver cuántas quedan completamente dentro del segmento y cuántas lo cubren por completo, con lo que tenemos una estimación por defecto y otra por exceso de la longitud del segmento cuya diferencia puede hacerse tan pequeña como queramos sin más que tomar divisiones suficientes.

Fagnano también mostró cómo dividir un cuadrante de lemniscata en 3 y 5 partes iguales.

Observemos que la aplicación inversa de la dada en el teorema 6.7 es algo más simple:

$$u(v) = \frac{2v\sqrt{1-v^4}}{1+v^4},$$

y en términos de esta aplicación el teorema nos dice cómo duplicar un arco de lemniscata (el arco que termina en $u(v)$ tiene el doble de longitud que el que termina en v). Euler se dio cuenta de que esta expresión era un caso particular de otra más general:

Teorema 6.9 *Para la parametrización de la lemniscata considerada en los teoremas anteriores, se cumple que*

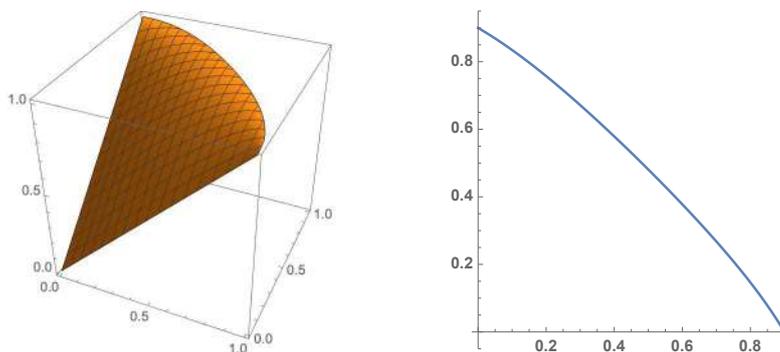
$$\int_0^{u_1} \frac{d\rho}{\sqrt{1-\rho^4}} + \int_0^{u_2} \frac{d\rho}{\sqrt{1-\rho^4}} = \int_0^v \frac{d\rho}{\sqrt{1-\rho^4}},$$

donde

$$v = \frac{u_1\sqrt{1-u_2^4} + u_2\sqrt{1-u_1^4}}{1+u_1^2u_2^2}.$$

Notemos que haciendo $u_1 = u_2$ en el teorema anterior se obtiene 6.7.

DEMOSTRACIÓN: La figura de la izquierda muestra la gráfica de la función $v(u_1, u_2)$:



De la propia definición se sigue que $v(u, 0) = v(0, u) = u$, lo que se traduce en la figura en que el corte con los planos de los ejes es una recta. Para cada altura v prefijada, la gráfica, es decir, el conjunto de puntos (u_1, u_2) para los que $v(u_1, u_2)$ toma el valor fijado, es una curva como la de la derecha, que une los puntos $(0, v)$ y $(v, 0)$. Así, si fijamos un valor para v , podemos considerar a u_2 como función de u_1 , es decir, llamamos $u_2(u_1)$ al valor que hace que $v(u_1, u_2(u_1))$ sea el valor v prefijado. Así, la función

$$f(u_1) = v(u_1, u_2(u_1)) = \frac{u_1 \sqrt{1 - u_2^4} + u_2 \sqrt{1 - u_1^4}}{1 + u_1^2 u_2^2}$$

es constante igual a v , luego su derivada es nula. Por la regla de la cadena será

$$f'(u_1) = \frac{\partial v}{\partial u_1} + \frac{\partial v}{\partial u_2} \frac{du_2}{du_1} = 0.$$

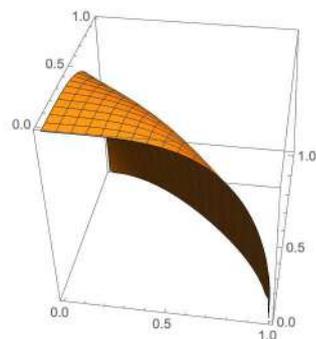
Vamos a calcular las derivadas:

$$\begin{aligned} \frac{\partial v}{\partial u_1} &= \frac{1}{(1 + u_1^2 u_2^2)^2} \left(\left(\sqrt{1 - u_2^4} - \frac{2u_2 u_1^3}{\sqrt{1 - u_1^4}} \right) (1 + u_1^2 u_2^2) \right. \\ &\quad \left. - \left(u_1 \sqrt{1 - u_2^4} + u_2 \sqrt{1 - u_1^4} \right) 2u_1 u_2^2 \right) = \\ &= \frac{\sqrt{1 - u_1^4} \sqrt{1 - u_2^4} (1 - u_1^2 u_2^2) - 2(u_1 u_2)(u_1^2 + u_2^2)}{(1 + u_1^2 u_2^2)^2 \sqrt{1 - u_1^4}} \end{aligned}$$

Por simetría:

$$\frac{\partial v}{\partial u_2} = \frac{\sqrt{1 - u_1^4} \sqrt{1 - u_2^4} (1 - u_1^2 u_2^2) - 2(u_1 u_2)(u_1^2 + u_2^2)}{(1 + u_1^2 u_2^2)^2 \sqrt{1 - u_2^4}}.$$

La gráfica siguiente muestra el numerador común de ambas derivadas, y vemos —se puede comprobar algebraicamente— que sólo se anula sobre la curva correspondiente a $v = 1$.



Equivalentemente, si $v < 1$, los numeradores no se anulan, luego en la igualdad de la regla de la cadena podemos simplificarlos, al igual que el factor común del denominador, y la igualdad se reduce a

$$\frac{1}{\sqrt{1-u_1^4}} + \frac{1}{\sqrt{1-u_2^4}} \frac{du_2}{du_1} = 0.$$

Integrando para u_1 entre 0 y un valor $u_1 < 1$ (y aplicando el teorema de cambio de variable a la segunda parte) obtenemos

$$\int_0^{u_1} \frac{d\rho}{\sqrt{1-\rho^4}} + \int_v^{u_2(u_1)} \frac{d\rho}{\sqrt{1-\rho^4}} = 0,$$

donde hemos usado que $u_2(0) = v$, pues $v(0, u_2) = u_2$, luego el valor de u_2 necesario para conseguir el valor v cuando $u_1 = 0$ es $u_2(0) = v$. Por lo tanto, llamando $u_2 = u_2(u_1)$,

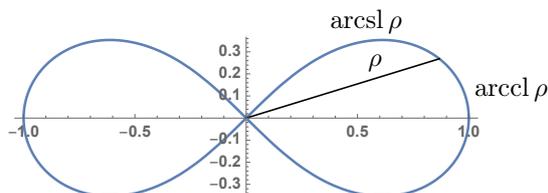
$$\begin{aligned} \int_0^{u_1} \frac{d\rho}{\sqrt{1-\rho^4}} + \int_0^{u_2} \frac{d\rho}{\sqrt{1-\rho^4}} &= \int_0^{u_1} \frac{d\rho}{\sqrt{1-\rho^4}} + \int_0^v \frac{d\rho}{\sqrt{1-\rho^4}} + \int_v^{u_2} \frac{d\rho}{\sqrt{1-\rho^4}} \\ &= \int_0^v \frac{d\rho}{\sqrt{1-\rho^4}}. \end{aligned} \quad \blacksquare$$

En la sección siguiente daremos pruebas más elegantes de los teoremas que acabamos de demostrar.

6.3 Las funciones elípticas lemniscáticas

Tal y como hemos explicado en la introducción, Gauss se dio cuenta de que los teoremas de Fagnano y Euler sobre la lemniscata podían expresarse de forma mucho más natural en términos de la función inversa de la dada por la integral lemniscática. Concretamente, Gauss definió el *arco seno* y el *arco coseno lemniscáticos* como

$$\operatorname{arcsl} \rho = \int_0^\rho \frac{dx}{\sqrt{1-x^4}}, \quad \operatorname{arcl} \rho = \int_\rho^1 \frac{dx}{\sqrt{1-x^4}},$$



de modo que, llamando

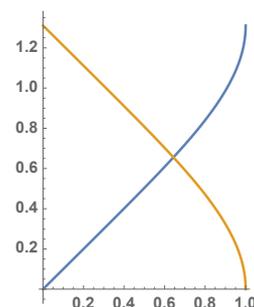
$$\varpi = 2 \int_0^1 \frac{dx}{\sqrt{1-x^4}} = \sqrt{2}K(1/2) = 2.622\dots$$

la longitud de la lemniscata completa es 2ϖ (o $2\varpi a$ si consideramos, en general, una lemniscata de semiamplitud a), y la de un cuarto de lemniscata es $\varpi/2$, y tenemos la relación

$$\operatorname{arcsl} \rho + \operatorname{arccl} \rho = \frac{\varpi}{2}. \quad (6.2)$$

La figura muestra las gráficas de las dos funciones. De su definición se sigue inmediatamente que

$$\operatorname{arcsl}, \operatorname{arccl} : [0, 1] \longrightarrow [0, \varpi/2],$$



así como que la primera es creciente y la segunda decreciente, y sus derivadas son

$$\frac{d \operatorname{arcsl} x}{dx} = \frac{1}{\sqrt{1-x^4}}, \quad \frac{d \operatorname{arccl} x}{dx} = -\frac{1}{\sqrt{1-x^4}}.$$

El teorema [ITAn 3.28] nos da que ambas tienen inversas continuas, a las que Gauss llamo *seno y coseno lemniscático*:

$$\operatorname{sl}, \operatorname{cl} : [0, \varpi/2] \longrightarrow [0, 1],$$

y el teorema de la función inversa 1.8 nos da que son derivables con derivadas

$$\frac{d \operatorname{sl} x}{dx} = \sqrt{1 - \operatorname{sl}^4 x}, \quad \frac{d \operatorname{cl} x}{dx} = -\sqrt{1 - \operatorname{cl}^4 x}.$$

Haciendo $\rho = \operatorname{cl} x$ en (6.2) obtenemos que $\operatorname{arcsl} \operatorname{cl} x + x = \varpi/2$, luego

$$\operatorname{cl} x = \operatorname{sl}(\varpi/2 - x).$$

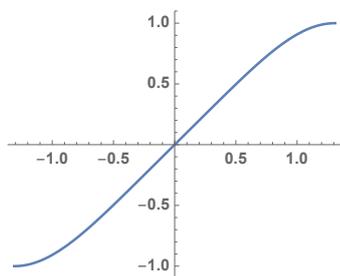
En realidad, la integral que define la función arcsl define una función

$$\operatorname{arcsl} : [-1, 1] \longrightarrow [-\varpi/2, \varpi/2]$$

que, por los mismos argumentos, tiene inversa derivable

$$\operatorname{sl} : [-\varpi/2, \varpi/2] \longrightarrow [-1, 1]$$

cuya gráfica es:



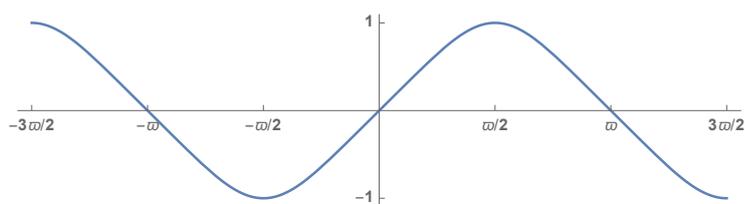
La gráfica y la analogía con la función seno sugiere extenderla a $[\varpi/2, 3\varpi/2]$ mediante

$$\operatorname{sl}(x + \varpi/2) = \operatorname{sl}(\varpi/2 - x)$$

y a $[-3\varpi/2, -\varpi/2]$ mediante

$$\operatorname{sl}(-x - \varpi/2) = \operatorname{sl}(x - \varpi/2),$$

con lo que obtenemos

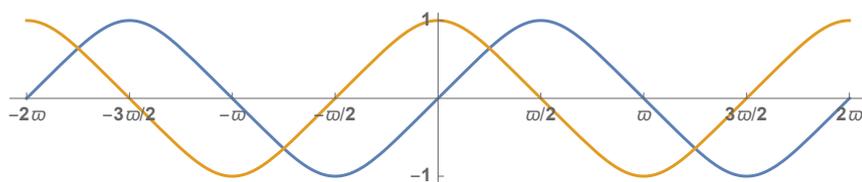


Notemos que esta definición equivale a que la gráfica de sl es simétrica respecto a las rectas verticales $x = \pm\varpi/2$. Además, tal y como muestra la gráfica, la función en $[\varpi/2, 3\varpi/2]$ toma los mismos valores que en $[-3\varpi/2, -\varpi/2]$. En efecto, para $0 \leq x \leq \varpi$,

$$\begin{aligned} \operatorname{sl}(-\varpi/2 - x) &= \operatorname{sl}(-\varpi/2 + x) = \operatorname{sl}(\varpi/2 - (\varpi - x)) \\ &= \operatorname{sl}(\varpi/2 + \varpi - x) = \operatorname{sl}(-\varpi/2 - x + 2\varpi). \end{aligned}$$

Esto se traduce en que podemos extender la función sl a una función periódica definida sobre toda la recta real con periodo 2ϖ , que será continua y derivable en todo su dominio.

Podemos hacer lo mismo con la función cl o, más fácilmente, extender su definición mediante la fórmula $\operatorname{cl}(x) = \operatorname{sl}(\varpi/2 - x)$. El resultado es:



De la construcción se sigue inmediatamente que

$$\operatorname{sl}(-x) = -\operatorname{sl} x$$

(se cumple en $[-\varpi/2, \varpi/2]$ por la definición como integral del arco seno lemniscático y se observa que se sigue cumpliendo en $[-\varpi/\varpi]$ por la simetría de la definición y a su vez en \mathbb{R} por la periodicidad).

Observemos ahora que la relación $\operatorname{sl}'(x) = \sqrt{1 - \operatorname{sl}^4(x)}$ sólo es válida en $[-\varpi/2, \varpi/2]$ (donde el seno lemniscático es creciente), pero la simetría de la extensión hace que en $[\varpi/2, 3\varpi/2]$ la relación correcta sea $\operatorname{sl}'(x) = -\sqrt{1 - \operatorname{sl}^4(x)}$, luego en todo el intervalo $[-\varpi/2, 3\varpi/2]$ se cumple

$$\operatorname{sl}'^2(x) = 1 - \operatorname{sl}^4 x, \quad \operatorname{cl}'^2(x) = 1 - \operatorname{cl}^4 x$$

(la segunda igualdad se deduce de la primera cambiando x por $\varpi/2 - x$) y, como las funciones tienen periodo 2ϖ , ambas igualdades son válidas para todo número real x . Volviendo a derivar:

$$2 \operatorname{sl}' x \operatorname{sl}'' x = -4 \operatorname{sl}^3 x \operatorname{sl}' x, \quad 2 \operatorname{cl}' x \operatorname{cl}'' x = -4 \operatorname{cl}^3 x \operatorname{cl}' x,$$

luego

$$\operatorname{sl}'' x = -2 \operatorname{sl}^3 x, \quad \operatorname{cl}'' x = -2 \operatorname{cl}^3 x.$$

Ahora probamos un teorema de adición:

Teorema 6.10 *Para todo par de números reales x, β se cumple:*

$$\operatorname{sl}(x + \beta) = \frac{\operatorname{sl} x \operatorname{sl}' y + \operatorname{sl} y \operatorname{sl}' x}{1 + \operatorname{sl}^2 x \operatorname{sl}^2 y}.$$

DEMOSTRACIÓN: Fijemos un número real γ y llamemos

$$s_1(x) = \operatorname{sl} x, \quad s_2(x) = \operatorname{sl}(\gamma - x).$$

Notemos que $s_2'(x) = -\operatorname{sl}'(\gamma - x) = -2s_2^3(x)$, $s_2''(x) = \operatorname{sl}''(\gamma - x)$.

Basta probar que la expresión

$$f(x) = \frac{-s_1(x)s_2'(x) + s_2(x)s_1'(x)}{1 + s_1^2(x)s_2^2(x)}$$

es constante, pues en tal caso, haciendo $x = 0$ sale que la constante es $\operatorname{sl} \gamma$, y haciendo $\gamma = x + y$ obtenemos la fórmula del enunciado. Llamando $N(x)$ y $D(x)$ al numerador y al denominador, respectivamente, basta probar que

$$f'(x) = \frac{N'D - ND'}{D^2} = 0,$$

lo cual a su vez equivale a que $N'D = ND'$. En efecto:

$$\begin{aligned} N' &= -s_1' s_2' - s_1 s_2'' + s_2' s_2' + s_2 s_1'' = -s_1 s_2'' + s_2 s_1'' \\ &= 2s_1 s_2^3 - 2s_2 s_1^3 = 2s_1 s_2 (s_2^2 - s_1^2), \\ D' &= 2s_1 s_2' s_2^2 + 2s_1^2 s_2 s_2' = 2s_1 s_2 (s_1' s_2 + s_1 s_2'), \end{aligned}$$

luego la igualdad que queremos probar equivale a

$$2s_1s_2(s_2^2 - s_1^2)(1 + s_1^2s_2^2) = 2s_1s_2(s_1's_2 + s_1s_2')(-s_1s_2' + s_2s_1')$$

o también a

$$(s_2^2 - s_1^2)(1 + s_1^2s_2^2) = s_1'^2s_2^2 - s_1^2s_2'^2.$$

Usando que $s_1'^2 = 1 - s_1^4$, $s_2'^2 = 1 - s_2^4$ se llega inmediatamente a que la igualdad es correcta. ■

De aquí podemos deducir las propiedades básicas de las funciones lemniscáticas. Por ejemplo:

Teorema 6.11 *Se cumplen las relaciones siguientes:*

$$\operatorname{cl}^2(x) = \frac{1 - \operatorname{sl}^2x}{1 + \operatorname{sl}^2x}, \quad \operatorname{sl}^2(x) = \frac{1 - \operatorname{cl}^2x}{1 + \operatorname{cl}^2x}, \quad (6.3)$$

$$\operatorname{cl}^2x + \operatorname{sl}^2x + \operatorname{cl}^2x\operatorname{sl}^2x = 1. \quad (6.4)$$

$$\frac{d\operatorname{sl}x}{dx} = \operatorname{cl}x(1 + \operatorname{sl}^2x), \quad \frac{d\operatorname{cl}x}{dx} = -\operatorname{sl}x(1 + \operatorname{cl}^2x). \quad (6.5)$$

DEMOSTRACIÓN: Por el teorema anterior, como $\operatorname{sl}(\varpi/2) = 1$, $\operatorname{sl}'(\varpi/2) = 0$:

$$\operatorname{cl}(x) = \operatorname{sl}(\varpi/2 - x) = \frac{\operatorname{sl}(\varpi/2)\operatorname{sl}'(-x) + \operatorname{sl}x\operatorname{sl}'(\varpi/2)}{1 + \operatorname{sl}^2(\varpi/2)\operatorname{sl}^2x} = \pm \frac{\sqrt{1 - \operatorname{sl}^4x}}{1 + \operatorname{sl}^2x},$$

luego

$$\operatorname{cl}^2(x) = \frac{(1 - \operatorname{sl}^2x)(1 + \operatorname{sl}^2x)}{(1 + \operatorname{sl}^2x)^2} = \frac{1 - \operatorname{sl}^2x}{1 + \operatorname{sl}^2x}.$$

Operando obtenemos la relación fundamental (6.4). Sumando 1 a sus dos miembros se convierte en

$$(1 + \operatorname{cl}^2x)(1 + \operatorname{sl}^2x) = 2,$$

de donde podemos despejar (6.3). Finalmente,

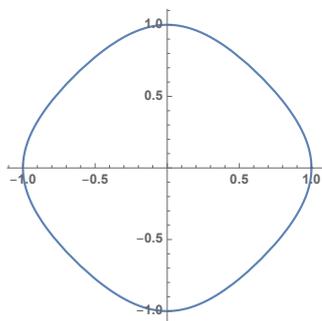
$$\operatorname{sl}'^2x = 1 - \operatorname{sl}^4x = (1 + \operatorname{sl}^2x)(1 - \operatorname{sl}^2x) = (1 + \operatorname{sl}^2x)^2 \frac{1 - \operatorname{sl}^2x}{1 + \operatorname{sl}^2x} = (1 + \operatorname{sl}^2x)^2 \operatorname{cl}^2x,$$

y basta tomar la raíz cuadrada observando que el coseno lemniscático es positivo donde el seno lemniscático es creciente, es decir, donde su derivada es positiva. Igualmente se deduce la segunda fórmula de (6.5), salvo que en este caso es necesario un signo negativo, porque el seno lemniscático es positivo donde el coseno lemniscático es decreciente. ■

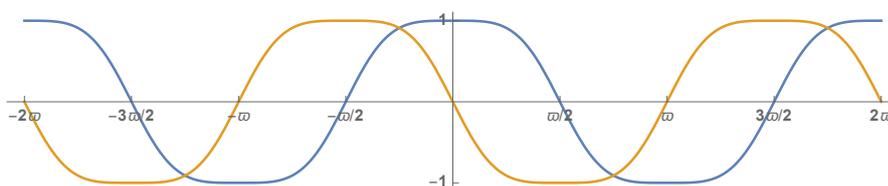
La relación fundamental (6.4) muestra que las funciones cl y sl no parametrizan la lemniscata, sino la curva de ecuación

$$x^2 + y^2 + x^2y^2 = 1$$

cuya gráfica es:



He aquí las gráficas de las derivadas de las funciones lemniscáticas:



Del teorema 6.10 se sigue inmediatamente que $\text{sl}(x + \varpi) = -\text{sl } x$, o que

$$\text{sl}(x \pm \varpi/2) = \pm \frac{\text{sl}' x}{1 + \text{sl}^2 x} = \pm \text{cl } x, \quad \text{cl}(x \pm \varpi/2) = \mp \text{sl } x. \quad (6.6)$$

(La segunda igualdad se deduce de la primera sustituyendo x por $x \mp \varpi/2$.)

Si en 6.10 sustituimos las derivadas por su expresión según 6.11 queda una expresión más complicada que la dada por el teorema siguiente:

Teorema 6.12 *Se cumplen las identidades siguientes:*

$$\text{sl}(x + y) = \frac{\text{sl } x \text{cl } y + \text{sl } y \text{cl } x}{1 - \text{sl } x \text{cl } x \text{sl } y \text{cl } y}, \quad \text{cl}(x + y) = \frac{\text{cl } x \text{cl } y - \text{sl } x \text{sl } y}{1 + \text{sl } x \text{cl } x \text{sl } y \text{cl } y}.$$

La demostración la posponemos a la sección siguiente (página 381). En particular, de aquí obtenemos fórmulas de duplicación, triplicación, etc:

$$\text{sl } 2x = \frac{2 \text{sl } x \text{cl } x}{1 - \text{sl}^2 x \text{cl}^2 x}, \quad \text{cl } 2x = \frac{\text{cl}^2 x - \text{sl}^2 x}{1 + \text{sl}^2 x \text{cl}^2 x}.$$

Para calcular $\text{sl } 3x$ conviene expresar $\text{sl } 2x$ y $\text{cl } 2x$ en términos casi exclusivamente de $\text{sl } x$, usando (6.3):

$$\text{sl } 2x = \frac{2 \text{sl } x (1 + \text{sl}^2 x) \text{cl } x}{1 + \text{sl}^4 x}, \quad \text{cl } 2x = \frac{\text{sl}^4 x + 2 \text{sl}^2 x - 1}{\text{sl}^4 x - 2 \text{sl}^2 x - 1}.$$

Así:

$$\text{sl } 3x = \frac{\text{sl } 2x \text{cl } x + \text{sl } x \text{cl } 2x}{1 - \text{sl } 2x \text{cl } 2x \text{sl } x \text{cl } x} = \frac{\text{sl } 2x \text{cl } x + \text{sl } x \text{cl } 2x}{1 - \text{sl } 2x \text{cl } 2x \text{sl } x \text{cl } x}$$

Antes sustituir $\operatorname{sl} 2x$ observamos que $\operatorname{cl}^2 x(1 + \operatorname{sl}^2 x) = 1 - \operatorname{sl}^2 x$, por lo que

$$\operatorname{sl} 3x = \frac{\frac{2 \operatorname{sl} x(1 - \operatorname{sl}^2 x)}{1 + \operatorname{sl}^4 x} + \operatorname{sl} x \frac{\operatorname{sl}^4 x + 2 \operatorname{sl}^2 x - 1}{\operatorname{sl}^4 x - 2 \operatorname{sl}^2 x - 1}}{1 - \frac{2 \operatorname{sl} x(1 - \operatorname{sl}^2 x)}{1 + \operatorname{sl}^4 x} \frac{\operatorname{sl}^4 x + 2 \operatorname{sl}^2 x - 1}{\operatorname{sl}^4 x - 2 \operatorname{sl}^2 x - 1} \operatorname{sl} x}.$$

Ahora operaciones rutinarias nos llevan a la expresión

$$\operatorname{sl} 3x = \frac{\operatorname{sl} x(\operatorname{sl}^8 x + 6 \operatorname{sl}^4 x - 3)}{3 \operatorname{sl}^8 x - 6 \operatorname{sl}^4 x - 1},$$

e igualmente se puede obtener que

$$\operatorname{cl} 3x = -\frac{\operatorname{cl} x(\operatorname{cl}^8 x + 6 \operatorname{cl}^4 x - 3)}{3 \operatorname{cl}^8 x - 6 \operatorname{cl}^4 x - 1}.$$

Para terminar observamos que los teoremas de Euler y Fagnano sobre la lemniscata son ahora inmediatos.

En efecto, si hacemos $u_1 = \operatorname{sl} x$ y $u_2 = \operatorname{sl} y$ en la suma de integrales del teorema 6.9, obtenemos la versión equivalente $x + y = \operatorname{arcsl} v$, luego $v = \operatorname{sl}(x + y)$, con lo que la fórmula para v del teorema 6.9 es simplemente la fórmula del teorema 6.10.

Similarmente, haciendo $u = \operatorname{sl} x$ en la igualdad de integrales del teorema 6.8, concluimos que $w = \operatorname{cl} x$, luego la fórmula para w de dicho teorema no es sino la dada por (6.3).

Más aún, ahora podemos deducir fácilmente el método de Fagnano para trisecar la lemniscata. Haciendo $x = \varpi/3$ en la fórmula para $\operatorname{sl} 3x$ obtenemos que

$$\frac{\operatorname{sl}(\varpi/3)(\operatorname{sl}^8(\varpi/3) + 6 \operatorname{sl}^4(\varpi/3) - 3)}{3 \operatorname{sl}^8(\varpi/3) - 6 \operatorname{sl}^4(\varpi/3) - 1} = \operatorname{sl} \varpi = 0,$$

con lo que

$$\operatorname{sl}^8 \frac{\varpi}{3} + 6 \operatorname{sl}^4 \frac{\varpi}{3} - 3 = 0,$$

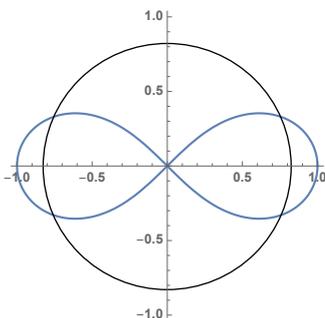
y así

$$\operatorname{sl}^4 \frac{\varpi}{3} = 2\sqrt{3} - 3, \quad \operatorname{sl} \frac{\varpi}{3} = \sqrt[4]{2\sqrt{3} - 3}.$$

Pero

$$\int_0^{\operatorname{sl}(\varpi/3)} \frac{d\rho}{\sqrt{1 - \rho^4}} = \frac{\varpi}{3}$$

es la longitud del arco de lemniscata determinado por la circunferencia de radio $\operatorname{sl}(\varpi/3)$, y este número es constructible con regla y compás:



Cada uno de los seis arcos en los que la circunferencia divide a la lemniscata mide $\varpi/3$, luego hemos dividido la lemniscata en 6 partes iguales. Juntándolas por pares tenemos la trisección.

6.4 Las funciones elípticas de Jacobi

Tal y como hemos explicado en la introducción, Jacobi invirtió la integral elíptica de primera especie:

$$F(\phi | m) = \int_0^\phi \frac{d\theta}{\sqrt{1 - m \operatorname{sen}^2 \theta}} = \int_0^{\operatorname{sen} \phi} \frac{1}{\sqrt{(1 - mt^2)(1 - t^2)}} dt.$$

Claramente, el primer integrando es positivo, por lo que la función $F(\phi | m)$ es estrictamente creciente en toda la recta real, Jacobi definió

$$\operatorname{am}(u | m),$$

como la inversa de $F(\phi | m)$, que recibe el nombre de *amplitud de Jacobi* de parámetro m . Por definición,

$$u = \int_0^{\operatorname{am}(u|m)} \frac{d\theta}{\sqrt{1 - m \operatorname{sen}^2 \theta}}.$$

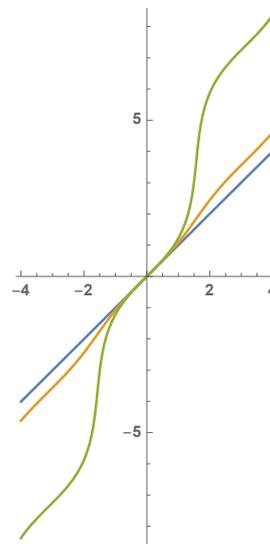
La figura muestra las amplitudes correspondientes a los parámetros $m = 0.01, 0.5, 0.99$. La de parámetro 0.01 es la más parecida a una recta, lo cual no es de extrañar, pues si extendemos la definición a $m = 0$ obtenemos claramente $\operatorname{am}(u | 0) = u$. Notemos que

$$\operatorname{am}(0 | m) = 0, \quad \operatorname{am}(K(m) | m) = \frac{\pi}{2}.$$

A su vez, las *funciones elípticas de Jacobi* se definen como

$$\operatorname{sn}(u | m) = \operatorname{sen} \operatorname{am}(u | m), \quad \operatorname{cn}(u | m) = \operatorname{cos} \operatorname{am}(u | m),$$

$$\operatorname{dn}(u | m) = \sqrt{1 - m \operatorname{sn}^2(u | m)}.$$

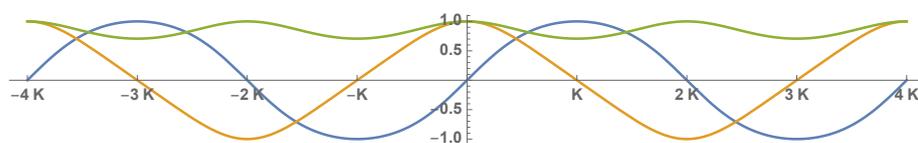


Así, por definición tenemos que

$$u = \int_0^{\operatorname{sn}(u|m)} \frac{1}{\sqrt{(1-mt^2)(1-t^2)}} dt. \quad (6.7)$$

De este modo, la función sn es la equivalente al seno lemniscático si que resulta de sustituir el polinomio $1-x^4$ en el integrando $1/\sqrt{1-x^4}$ por el polinomio $(1-mx^2)(1-x^2)$.

En lo sucesivo, y siempre que no se preste a confusión, suprimiremos la mención explícita del parámetro m . En particular, llamaremos $K = K(m)$. La figura muestra las gráficas para $m = 0.5$:



La que se parece al seno es sn y la que se parece al coseno es cn . De la definición se siguen inmediatamente las relaciones:

$$\operatorname{sn}^2(u) + \operatorname{cn}^2 u = 1, \quad \operatorname{dn}^2 u + m \operatorname{sn}^2 u = 1, \quad (6.8)$$

así como que

$$\operatorname{sn} 0 = 0, \quad \operatorname{cn} 0 = \operatorname{dn} 0 = 1, \quad \operatorname{sn} K = 1, \quad \operatorname{cn} K = 0, \quad \operatorname{dn} K = \sqrt{1-m}.$$

Vamos a probar que las tres funciones son periódicas. Para ello observamos que, aplicando los cambios de variable $\theta' = \pi - \theta$ y $\theta' = \theta - \pi$:

$$\begin{aligned} F(\phi + \pi) &= \int_0^{\phi+\pi} \frac{d\theta}{\sqrt{1-m \operatorname{sen}^2 \theta}} = \\ &= \int_0^{\pi/2} \frac{d\theta}{\sqrt{1-m \operatorname{sen}^2 \theta}} + \int_{\pi/2}^{\pi} \frac{d\theta}{\sqrt{1-m \operatorname{sen}^2 \theta}} + \int_{\pi}^{\phi+\pi} \frac{d\theta}{\sqrt{1-m \operatorname{sen}^2 \theta}} = \\ &= K - \int_{\pi/2}^0 \frac{d\theta}{\sqrt{1-m \operatorname{sen}^2 \theta}} + \int_0^{\phi} \frac{d\theta}{\sqrt{1-m \operatorname{sen}^2 \theta}} = 2K + F(\phi), \end{aligned}$$

luego, llamando $u = F(\phi)$, tenemos que $\operatorname{am}(u) + \pi = \operatorname{am}(u + 2K)$, luego

$$\operatorname{sn}(u + 2K) = -\operatorname{sn} u, \quad \operatorname{cn}(u + 2K) = -\operatorname{cn} u,$$

lo que a su vez implica que las funciones sn y cn tienen periodo $4K$, mientras que dn tiene periodo $2K$.

Para calcular sus derivadas empezamos observando que, si $u = F(\phi)$,

$$\frac{du}{d\phi} = \frac{1}{\sqrt{1-m \operatorname{sen}^2 \phi}}, \quad \frac{d\phi}{du} = \sqrt{1-m \operatorname{sen}^2 \phi} = \operatorname{dn} u.$$

De aquí se sigue a su vez que²

$$\frac{d \operatorname{sn} u}{du} = \operatorname{cn} u \operatorname{dn} u, \quad \frac{d \operatorname{cn} u}{du} = -\operatorname{sn} u \operatorname{dn} u,$$

$$\frac{d \operatorname{dn} u}{du} = -m \operatorname{sn} u \operatorname{cn} u.$$

A partir de estos hechos y de las propiedades de las funciones trigonométricas es fácil comprobar que las funciones elípticas de Jacobi se comportan en general como se aprecia en la gráfica anterior.

Observemos que las definiciones de las funciones elípticas de Jacobi son válidas igualmente para $m = 0, 1$, pero entonces dan lugar a funciones elementales: Es inmediato que $F(\phi | 0) = \phi$, de donde

$$\operatorname{sn}(u | 0) = \operatorname{sen} u, \quad \operatorname{cn}(u | 0) = \operatorname{cos} u, \quad \operatorname{dn}(u | 0) = 1, \quad K(0) = \pi/2.$$

Por otro lado, el cambio de variable $t = \operatorname{sen} \theta$ nos da que

$$F(\phi | 1) = \int_0^\phi \frac{d\theta}{\cos \theta} = \int_0^{\operatorname{sen} \phi} \frac{dt}{1-t^2} = \operatorname{artanh} \operatorname{sen} \phi,$$

luego

$$\operatorname{sn}(u | 1) = \operatorname{tanh} u, \quad \operatorname{cn}(u | 1) = \operatorname{dn}(u | 1) = \frac{1}{\cosh u}, \quad K(1) = \infty.$$

Las funciones de Jacobi también satisfacen fórmulas de adición:

Teorema 6.13 *Se cumplen las identidades siguientes:*

$$\begin{aligned} \operatorname{sn}(u+v) &= \frac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v + \operatorname{cn} u \operatorname{dn} u \operatorname{sn} v}{1 - m \operatorname{sn}^2 u \operatorname{sn}^2 v}, \\ \operatorname{cn}(u+v) &= \frac{\operatorname{cn} u \operatorname{cn} v - \operatorname{sn} u \operatorname{dn} u \operatorname{sn} v \operatorname{dn} v}{1 - m \operatorname{sn}^2 u \operatorname{sn}^2 v}, \\ \operatorname{dn}(u+v) &= \frac{\operatorname{dn} u \operatorname{dn} v - m \operatorname{sn} u \operatorname{cn} u \operatorname{sn} v \operatorname{cn} v}{1 - m \operatorname{sn}^2 u \operatorname{sn}^2 v}. \end{aligned}$$

DEMOSTRACIÓN: Fijemos un $\alpha \in \mathbb{R}$ arbitrario y consideremos las funciones $s_1(u) = \operatorname{sn} u$, $s_2(u) = \operatorname{sn}(\alpha - u)$. Entonces

$$s_1'(u) = \operatorname{cn} u \operatorname{dn} u, \quad s_2'(u) = -\operatorname{cn}(\alpha - u) \operatorname{dn}(\alpha - u).$$

Para demostrar la primera igualdad basta probar que la función

$$f(u) = \frac{s_1'(u)s_2(u) - s_2'(u)s_1(u)}{1 - m s_1^2(u)s_2^2(u)}$$

²Comparando con (6.5), vemos que la función dn desempeña el papel de las funciones $1 + \operatorname{sl}^2 x$ y $1 + \operatorname{cl}^2 x$ que aparecen en las fórmulas de las derivadas de las funciones sl y cl .

es constante, pues evaluando en $u = 0$ queda que la constante es $s_2(0) = \text{sn}(\alpha)$, luego tenemos la igualdad

$$\text{sn}(\alpha) = \frac{\text{sn } u \text{ cn}(\alpha - u) \text{ dn}(\alpha - u) + \text{cn } u \text{ dn}(\alpha - u) \text{ sn}(\alpha - u)}{1 - m \text{sn}^2 u \text{sn}^2(\alpha - u)}$$

y tomando $\alpha = u + v$ obtenemos la primera igualdad del enunciado.

Llamando $N = s'_1 s_2 - s'_2 s_1$ y $D = 1 - m s_1^2 s_2^2$, basta probar que

$$f'(u) = \frac{N'D - ND'}{D^2} = 0,$$

lo cual equivale a que $N'D = ND'$. Observemos que

$$s_1'^2 = (1 - s_1^2)(1 - m s_1^2), \quad s_2'^2 = (1 - s_2^2)(1 - m s_2^2), \quad (6.9)$$

y volviendo a derivar:

$$2s_1' s_1'' = -2s_1 s_1'(1 - m s_1^2) - (1 - s_1^2)2m s_1 s_1',$$

luego

$$s_1'' = -(1 + m)s_1 + 2m s_1^3.$$

Igualmente llegamos a que

$$s_2'' = -(1 + m)s_2 + 2m s_2^3.$$

Con esto ya podemos calcular

$$\begin{aligned} N' &= s_1'' s_2 + s_1' s_2' - s_2'' s_1 - s_2' s_1' = 2m s_1 s_2 (s_1^2 - s_2^2), \\ D' &= -2m s_1 s_1' s_2^2 - 2m s_1^2 s_2 s_2' = -2m s_1 s_2 (s_1' s_2 + s_1 s_2'), \end{aligned}$$

y queremos probar la igualdad

$$2m s_1 s_2 (s_1^2 - s_2^2)(1 - m s_1^2 s_2^2) = -2m s_1 s_2 (s_1' s_2 + s_1 s_2')(s_1' s_2 - s_2' s_1),$$

que equivale a

$$(s_1^2 - s_2^2)(1 - m s_1^2 s_2^2) = s_1^2 s_2'^2 - s_1'^2 s_2^2.$$

Sustituyendo (6.9) en el miembro derecho y operando se llega sin dificultad al miembro izquierdo.

Un cálculo rutinario muestra que si elevamos al cuadrado y sumamos los miembros derechos de las dos primeras fórmulas del enunciado el resultado es 1. Esto implica que la segunda fórmula se cumple salvo tal vez un signo, pero evaluándola en $u = 0$ se ve que el signo tiene que ser el positivo.

La tercera fórmula se prueba igualmente, ahora multiplicando por m el cuadrado del miembro izquierdo de la primera igualdad. ■

Por ejemplo, observando la gráfica anterior vemos que la gráfica de la función cn no es una traslación de la gráfica de sn , al contrario de lo que sucede con las

funciones cos y sen usuales, e incluso con las funciones lemniscáticas cl y sl. El teorema anterior nos da la relación correcta:

$$\operatorname{sn}(u+K) = \frac{\operatorname{cn} u}{\operatorname{dn} u}, \quad \operatorname{cn}(u+K) = -\sqrt{1-m} \frac{\operatorname{sn} u}{\operatorname{dn} u}, \quad \operatorname{dn}(u+K) = \frac{\sqrt{1-m}}{\operatorname{dn} u}.$$

Las funciones lemniscáticas se pueden expresar en términos de las funciones de Jacobi:

Teorema 6.14 Para $m = 1/2$ se cumple:

$$\operatorname{sl} u = \frac{\sqrt{2}}{2} \frac{\operatorname{sn}(\sqrt{2}u)}{\operatorname{dn}(\sqrt{2}u)}, \quad \operatorname{cl} u = \operatorname{cn}(\sqrt{2}u).$$

DEMOSTRACIÓN: Llamemos $y(u) = \operatorname{cn}(\sqrt{2}u)$, que es una función periódica de periodo $4K(1/2)/\sqrt{2} = 2\sqrt{2}K(1/2) = 2\varpi$, al igual que cl. Por lo tanto, basta probar que coinciden en $[-\varpi, \varpi]$, pero ambas cumplen $y(-u) = y(u)$, $\operatorname{cl}(-u) = \operatorname{cl}(u)$, luego basta probar que coinciden en $[0, \varpi]$ y, como además, por las fórmulas de adición,

$$y(2K-u) = -y(u), \quad \operatorname{cl}(\varpi-u) = -\operatorname{cl}(u),$$

basta probar que coinciden en $[0, \varpi/2] = [0, K]$, donde ambas son positivas. Observamos que

$$y'(u) = -\sqrt{2} \operatorname{sn}(\sqrt{2}u) \operatorname{dn}(\sqrt{2}u),$$

luego

$$\begin{aligned} y'^2(u) &= 2 \operatorname{sn}^2(\sqrt{2}u) \operatorname{dn}^2(\sqrt{2}u) = 2(1 - \operatorname{cn}^2(\sqrt{2}u))(1 - (1/2) \operatorname{sn}^2(\sqrt{2}u)) \\ &= (1 - \operatorname{cn}^2(\sqrt{2}u))(2 - \operatorname{sn}^2(\sqrt{2}u)) = (1 - \operatorname{cn}^2(\sqrt{2}u))(1 + \operatorname{cn}^2(\sqrt{2}u)) \\ &= 1 - \operatorname{cn}^4(\sqrt{2}u) = 1 - y(u)^4. \end{aligned}$$

Así pues, $y'(u) = \sqrt{1-y^4(u)}$, luego aplicando el cambio de variable $y(u)$ vemos que

$$\operatorname{arcsl} y(u) = \int_0^{y(u)} \frac{dr}{\sqrt{1-r^4}} = \int_{\varpi/2}^u du = u - \frac{\varpi}{2},$$

luego

$$y(u) = \operatorname{sl}(u - \varpi/2) = \operatorname{sl}(\varpi/2 - u) = \operatorname{cl} u.$$

Ahora usamos (6.3) y (6.8):

$$\operatorname{sl}^2 u = \frac{1 - \operatorname{cn}^2(\sqrt{2}u)}{1 + \operatorname{cn}^2(\sqrt{2}u)} = \frac{\operatorname{sn}^2(\sqrt{2}u)}{2 \operatorname{dn}^2(\sqrt{2}u)},$$

luego

$$\operatorname{sl} u = \pm \frac{\sqrt{2}}{2} \frac{\operatorname{sn}(\sqrt{2}u)}{\operatorname{dn}(\sqrt{2}u)},$$

pero las funciones sl y sn son positivas en los mismos intervalos, luego el signo correcto es el positivo. ■

Alternativamente, podríamos haber usado este resultado para definir las funciones lemniscáticas y demostrar a partir de aquí sus propiedades. Por ejemplo, ahora es fácil demostrar el teorema 6.12:

DEMOSTRACIÓN (de 6.12): Por los teoremas 6.14 y 6.13 tenemos que

$$\begin{aligned} \operatorname{sl}(x+y) &= \frac{(1/2) \operatorname{sn}(\sqrt{2}x + \sqrt{2}y)}{\operatorname{dn}(\sqrt{2}x + \sqrt{2}y)} = \\ &= \frac{(1/2) \operatorname{sn}(\sqrt{2}x) \operatorname{cn}(\sqrt{2}y) \operatorname{dn}(\sqrt{2}y) - \operatorname{cn}(\sqrt{2}x) \operatorname{dn}(\sqrt{2}x) \operatorname{sn}(\sqrt{2}y)}{\operatorname{dn}(\sqrt{2}x) \operatorname{dn}(\sqrt{2}y) - (1/2) \operatorname{sn}(\sqrt{2}x) \operatorname{cn}(\sqrt{2}x) \operatorname{sn}(\sqrt{2}y) \operatorname{cn}(\sqrt{2}y)}. \end{aligned}$$

Dividiendo numerador y denominador entre $\operatorname{dn}(\sqrt{2}x) \operatorname{dn}(\sqrt{2}y)$ se obtiene la primera fórmula del enunciado. Para la segunda usamos (6.6):

$$\begin{aligned} \operatorname{cl}(x+y) &= \operatorname{sl}(x + (y + \varpi/2)) = \frac{\operatorname{sl} x \operatorname{cl}(y + \varpi/2) + \operatorname{sl}(y + \varpi/2) \operatorname{cl} x}{1 - \operatorname{sl} x \operatorname{cl} x \operatorname{sl}(y + \varpi/2) \operatorname{cl}(y + \varpi/2)} \\ &= \frac{-\operatorname{sl} x \operatorname{sl} y + \operatorname{cl} x \operatorname{cl} y}{1 + \operatorname{sl} x \operatorname{cl} x \operatorname{cl} y \operatorname{sl} y}. \end{aligned}$$

■

Como primera aplicación de las funciones de Jacobi vamos a probar un teorema de adición para la integral elíptica de segunda especie. En primer lugar observemos que, en la expresión

$$E(\theta | m) = \int_0^\theta \sqrt{1 - m \operatorname{sen}^2 \phi} d\phi$$

podemos hacer el cambio $\phi = \operatorname{am}(u)$, y así $\operatorname{sen} \phi = \operatorname{sn} u$, $\operatorname{cos} \phi = \operatorname{cn} u$, luego

$$\operatorname{cos} \phi d\phi = \operatorname{cn} u \operatorname{dn} u du, \quad d\phi = \operatorname{dn} u du$$

y la integral se convierte en

$$E(u : m) = \int_0^u \operatorname{dn}^2 u du.$$

Es tomando u como argumento como el teorema de adición tiene su expresión más natural:

Teorema 6.15

$$E(u_1 + u_2) = E(u_1) + E(u_2) - m \operatorname{sn} u_1 \operatorname{sn} u_2 \operatorname{sn}(u_1 + u_2).$$

DEMOSTRACIÓN: Según 6.13:

$$\begin{aligned} \operatorname{dn}(x+y) + \operatorname{dn}(x-y) &= \frac{2 \operatorname{dn} x \operatorname{dn} y}{1 - m \operatorname{sn}^2 x \operatorname{sn}^2 y}, \\ \operatorname{dn}(x+y) - \operatorname{dn}(x-y) &= \frac{-2m \operatorname{sn} x \operatorname{cn} x \operatorname{sn} y \operatorname{cn} y}{1 - m \operatorname{sn}^2 x \operatorname{sn}^2 y}, \end{aligned}$$

y multiplicando ambas expresiones queda

$$\operatorname{dn}^2(x+y) - \operatorname{dn}^2(x-y) = \frac{-4m \operatorname{sn} x \operatorname{sn} y \operatorname{cn} x \operatorname{cn} y \operatorname{dn} x \operatorname{dn} y}{(1 - m \operatorname{sn}^2 x \operatorname{sn}^2 y)^2}.$$

Ahora observamos que

$$\frac{d \operatorname{dn}^2 u}{du} = E(u : m),$$

luego

$$\frac{d \operatorname{dn}^2(x+y)}{dy} = E(x+y : m), \quad \frac{d \operatorname{dn}^2(x-y)}{dy} = -E(x+y : m),$$

luego integrando la igualdad que hemos obtenido resulta que

$$E(x+y) + E(x-y) = C_x + \frac{2 \operatorname{cn} x \operatorname{dn} x}{\operatorname{sn} x} \int \frac{-2m \operatorname{sn}^2 x \operatorname{sn} y \operatorname{cn} y \operatorname{dn} y}{(1 - m \operatorname{sn}^2 x \operatorname{sn}^2 y)^2} dy.$$

En la integral que queda, el numerador es la derivada de la base del denominador, luego

$$E(x+y) + E(x-y) = C_x - \frac{2 \operatorname{sn} x \operatorname{cn} x \operatorname{dn} x}{\operatorname{sn}^2 x (1 - m \operatorname{sn}^2 x \operatorname{sn}^2 y)}.$$

Para $y = x$ queda

$$E(2x) = C_x - \frac{2 \operatorname{sn} x \operatorname{cn} x \operatorname{dn} x}{\operatorname{sn}^2 x (1 - m \operatorname{sn}^4 x)},$$

y restando las dos expresiones eliminamos la constante de integración:

$$E(x+y) + E(x-y) - E(2x) = m \frac{2 \operatorname{sn} x \operatorname{cn} x \operatorname{dn} x}{1 - m \operatorname{sn}^4 x} \frac{\operatorname{sn}^2 x - \operatorname{sn}^2 y}{1 - m \operatorname{sn}^2 x \operatorname{sn}^2 y}.$$

De 6.13 se sigue inmediatamente que la primera fracción es $\operatorname{sn} 2x$, y es fácil ver que la segunda es $\operatorname{sn}(x+y) \operatorname{sn}(x-y)$, con lo que llegamos a que

$$E(x+y) + E(x-y) - E(2x) = m \operatorname{sn} 2x \operatorname{sn}(x+y) \operatorname{sn}(x-y).$$

Ahora, dados dos números u_1 y u_2 , siempre podemos tomar x, y de manera que $u_1 = x+y$, $u_2 = x-y$ (luego $u_1 + u_2 = 2x$), y así

$$E(u_1) + E(u_2) - E(u_1 + u_2) = m \operatorname{sn}(u_1 + u_2) \operatorname{sn} u_1 \operatorname{sn} u_2. \quad \blacksquare$$

Ahora vamos a ver que este teorema es el trasfondo del cálculo que hizo Fagnano sobre los arcos de una elipse:

Teorema 6.16 *Consideremos una elipse*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

de excentricidad $\epsilon = \sqrt{1 - b^2/a^2}$ y tomemos en ella dos puntos

$$P = (x, y) = (a \cos \phi, b \operatorname{sen} \phi), \quad P' = (x', y') = (a \cos \phi', b \operatorname{sen} \phi')$$

tales que

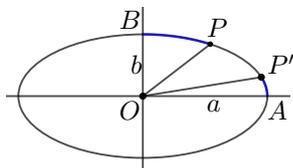
$$\tan \phi \tan \phi' = \frac{b}{a}.$$

Entonces

$$\widehat{BP} - \widehat{P'A} = \epsilon^2 \frac{xx'}{a}.$$

En particular, si F es el punto que cumple $F = F'$, se cumple que

$$\widehat{BF} - \widehat{FA} = a - b.$$



Vamos a considerar las funciones elípticas de Jacobi de módulo $m = \epsilon^2$, de modo que $b/a = \sqrt{1 - m}$.

En este punto hay que destacar que, para que la integral que nos da la longitud de un arco de elipse se traduzca fácilmente en una integral elíptica necesitamos parametrizar la elipse en la forma

$$r(\phi) = (a \operatorname{sen} \phi, b \cos \phi),$$

que no es la habitual considerada en el enunciado. Si ϕ es el parámetro de un punto según la parametrización del enunciado, en la que consideramos ahora su parámetro será $\phi^* = \pi/2 - \phi$, por lo que $\tan \phi = 1/\tan \phi^*$, y en lo único en lo que se ve afectado el enunciado es que, respecto a esta parametrización, la condición sobre las tangentes es

$$\tan \phi \tan \phi' = \frac{a}{b}.$$

Con la parametrización invertida tenemos que

$$aE(\phi | m) = a \int_0^\phi \sqrt{1 - \operatorname{sen}^2 \theta} d\theta$$

es la longitud del arco comprendido entre el punto B y el punto P de amplitud ϕ . Ahora hacemos el cambio de parámetro $\phi = \operatorname{am}(u | m)$, con lo que pasamos a la parametrización $(x, y) = (a \operatorname{sn} u, b \operatorname{cn} u)$, y la longitud del arco comprendido entre B y el punto $P = (x, y)$ de parámetro u viene dada por $aE(u : m)$.

Tomemos ahora $u' = K - u$, que se corresponde con otro punto de la elipse $P' = (x', y') = (a \operatorname{sn} u', b \operatorname{cn} u')$. El teorema anterior nos da la relación

$$E(K) = E(u) + E(u') - \epsilon^2 \operatorname{sn} u \operatorname{sn} u' = E(u) + E(u') - \epsilon^2 \frac{xx'}{a^2},$$

Multiplicando por a queda

$$aE(u) + aE(u') - L = \epsilon^2 \frac{xx'}{a},$$

donde $L = E(K)$ es la longitud del arco \widehat{AB} , luego la igualdad equivale a

$$\widehat{BP} - \widehat{P'A} = \epsilon^2 \frac{xx'}{a}.$$

Hemos probado que esto se cumple para cualquier par de puntos P y P' correspondientes a parámetros $u + u' = K$. Por 6.13, esto implica que

$$\operatorname{sn} u' = \frac{\operatorname{cn} u}{\operatorname{dn} u}, \quad \operatorname{cn} u' = \sqrt{1-m} \frac{\operatorname{sn} u}{\operatorname{dn} u} = \frac{b}{a} \frac{\operatorname{sn} u}{\operatorname{dn} u}.$$

Por lo tanto,

$$\frac{a}{b} \tan \phi = \frac{x}{y} = \frac{a \operatorname{sn} u}{b \operatorname{cn} u}, \quad \frac{a}{b} \tan \phi' = \frac{x'}{y'} = \frac{a \operatorname{sn} u'}{b \operatorname{cn} u'} = \frac{a^2 \operatorname{cn} u}{b^2 \operatorname{sn} u},$$

luego

$$\tan \phi \tan \phi' = \frac{a}{b}.$$

El punto de Fagnano cumple

$$\frac{a}{b} = \tan^2 \phi = \frac{b^2 x^2}{a^2 y^2},$$

luego $a^3 y^2 = b^3 x^2$. Sustituyendo en la ecuación de la elipse resulta que

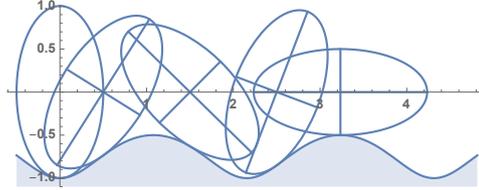
$$F = \left(a \sqrt{\frac{a}{a+b}}, b \sqrt{\frac{b}{a+b}} \right)$$

y el miembro derecho de la fórmula del enunciado se reduce a

$$\frac{\epsilon^2 x^2}{a} = \left(1 - \frac{b^2}{a^2}\right) a \frac{a}{a+b} = a - b.$$

■

Ejemplo: Ruedas elípticas En la página 132 vimos cuál es la curva sobre la cual una rueda cuadrada se mueve de modo que su eje sigue una trayectoria horizontal. Ahora vamos a determinar la curva que consigue lo mismo con una rueda elíptica:



Consideramos, concretamente, una elipse de semiejes a y b , luego de excentricidad $\epsilon = \sqrt{1 - b^2/a^2}$, y suponemos que parte de la posición “vertical” que muestra la figura a la izquierda, con su centro en el origen de coordenadas. Buscamos la función $y(x)$ cuya gráfica es la forma del suelo. En particular, tiene que ser $y(0) = -a$. A partir de ahí, $y(x)$ tiene que ir subiendo hasta llegar a un punto de altura $-b$, en el que se apoyará la elipse cuando haya girado 90° y se haya puesto horizontal. Basta describir $y(x)$ en el intervalo de tiempo en que pasa de $-a$ a $-b$, pues es claro que a partir de ahí debe decrecer de forma simétrica hasta volver a tomar el valor $-a$, y a partir de ese punto se prolonga periódicamente. En particular, en el intervalo que nos interesa, es $y'(x) > 0$.

Consideremos el instante en que la elipse ha girado un ángulo θ . Por (1.3) sabemos que la subnormal de f en la posición x del centro de la elipse es

$$S = -y \frac{dy}{dx}.$$

(donde le hemos antepuesto un signo negativo para que sea $S > 0$).

Ahora vamos a ver la situación “desde el punto de vista de la elipse”. Consideramos elipse de ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

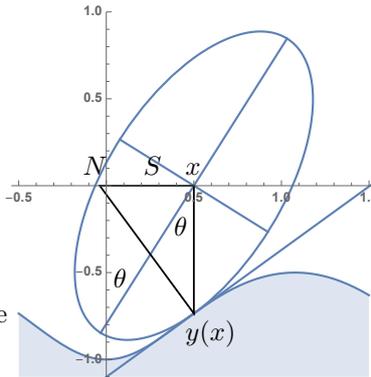
de modo que las coordenadas polares de sus puntos cumplen

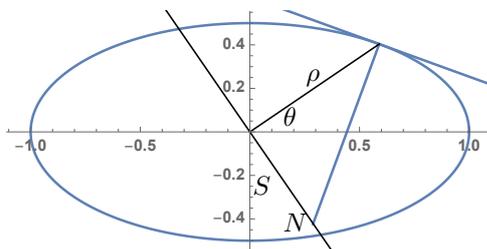
$$\rho^2 \left(\frac{\cos^2 \theta}{a^2} + \frac{\sin^2 \theta}{b^2} \right) = 1,$$

luego la ecuación en coordenadas polares de la elipse es

$$\rho = \left(\frac{\cos^2 \theta}{a^2} + \frac{\sin^2 \theta}{b^2} \right)^{-1/2}.$$

Y ahora sólo tenemos que observar que S es también la subnormal polar de la elipse $S = -\rho'$:





Vamos a calcular S a partir de la elipse, para lo cual partimos de

$$\rho^{-2} = \frac{\cos^2 \theta}{a^2} + \frac{\sin^2 \theta}{b^2}. \quad (6.10)$$

Al derivar queda:

$$-2\rho^{-3} \frac{d\rho}{d\theta} = \left(\frac{1}{b^2} - \frac{1}{a^2} \right) 2 \cos \theta \sin \theta.$$

Por otro lado, si en (6.10) sustituimos $\cos^2 \theta = 1 - \sin^2 \theta$ o $\sin^2 \theta = 1 - \cos^2 \theta$, obtenemos:

$$\frac{1}{\rho^2} - \frac{1}{a^2} = \left(\frac{1}{b^2} - \frac{1}{a^2} \right) \sin^2 \theta, \quad (6.11)$$

$$\frac{1}{b^2} - \frac{1}{\rho^2} = \left(\frac{1}{b^2} - \frac{1}{a^2} \right) \cos^2 \theta,$$

luego

$$\begin{aligned} -\frac{2}{\rho^3} \frac{d\rho}{d\theta} &= 2\sqrt{\frac{1}{b^2} - \frac{1}{a^2}} \cos \theta \sqrt{\frac{1}{b^2} - \frac{1}{a^2}} \sin \theta = 2\sqrt{\left(\frac{1}{b^2} - \frac{1}{a^2}\right) \left(\frac{1}{\rho^2} - \frac{1}{a^2}\right)} \\ &= \frac{\sqrt{(a^2 - \rho^2)(\rho^2 - b^2)}}{\rho^2 ab}, \end{aligned}$$

luego

$$S = -\frac{d\rho}{d\theta} = \frac{\rho \sqrt{(a^2 - \rho^2)(\rho^2 - b^2)}}{ab}.$$

Aquí ρ es la distancia del centro de la elipse al punto de contacto con el suelo, que, en términos de la función $y(x)$, es precisamente $-y(x)$, luego igualando las dos expresiones para S concluimos que

$$y \frac{dy}{dx} = \frac{y \sqrt{(a^2 - y^2)(y^2 - b^2)}}{ab},$$

y así tenemos una ecuación diferencial para la función $y(x)$ que buscamos.

Separamos las variables:

$$dx = \frac{ab}{\sqrt{(a^2 - y^2)(y^2 - b^2)}} dy,$$

de donde, teniendo en cuenta la condición inicial $y(0) = -a$,

$$x = \int_{-a}^y \frac{ab}{\sqrt{(a^2 - t^2)(t^2 - b^2)}} dt.$$

Hacemos el cambio de variable

$$s = \sqrt{\frac{a^2 - t^2}{a^2 - b^2}}, \quad t^2 = a^2 - (a^2 - b^2)s^2,$$

$$a^2 - t^2 = (a^2 - b^2)s^2, \quad t^2 - b^2 = (a^2 - b^2)(1 - s^2),$$

$$2t dt = -2(a^2 - b^2)s ds, \quad dt = \frac{-(a^2 - b^2)s}{-\sqrt{a^2 - (a^2 - b^2)s^2}} ds = \frac{(a^2 - b^2)s}{\sqrt{1 - \epsilon^2 s^2}} ds.$$

Así:

$$x = b \int_0^{\tilde{y}} \frac{1}{\sqrt{(1 - s^2)(1 - \epsilon^2 s^2)}} ds$$

donde

$$\tilde{y} = \sqrt{\frac{a^2 - y^2}{a^2 - b^2}}.$$

Ahora bien, según (6.7), esto equivale a que

$$\sqrt{\frac{a^2 - y^2}{a^2 - b^2}} = \tilde{y} = \operatorname{sn}(x/b | \epsilon^2).$$

Despejando llegamos a que

$$y^2 = a^2(1 - \epsilon^2 \operatorname{sn}(x/b | \epsilon^2)),$$

luego la curva que buscábamos es, esencialmente, la función dn:

$$y = -a \operatorname{dn}(x/b | \epsilon^2).$$

En principio hemos probado esto para $0 \leq x \leq K$, es decir, mientras $y' \geq 0$, pero la función que hemos obtenido desciende simétricamente entre K y $2K$, al igual que tiene que sucederle a y , y tiene periodo $2K$, luego concluimos que la igualdad anterior es válida para todo x .

Teniendo en cuenta que $\rho = -y(x)$, la fórmula (6.11) nos da el ángulo θ que corresponde a cada posición del centro x de la elipse:

$$\operatorname{sen} \theta = \pm \frac{b\sqrt{a^2 - y^2}}{ya\epsilon}.$$

■

6.5 El péndulo simple

En la sección 4.2 hemos visto que la amplitud θ de un péndulo simple satisface la ecuación diferencial (4.3):

$$\theta'' = -\frac{g}{l} \operatorname{sen} \theta,$$

donde l es la longitud de la cuerda y θ el ángulo respecto de la posición vertical en la que el péndulo está en reposo. En lugar de resolverla, allí pasamos a la aproximación que resulta de considerar $\operatorname{sen} \theta \approx \theta$, que es válida para amplitudes pequeñas. Ahora estamos en condiciones de resolver la ecuación de forma exacta. Como no vamos a restringirnos a amplitudes pequeñas, conviene considerar que la masa que cuelga está sujeta, no a una cuerda, sino más bien a una varilla rígida de masa despreciable. La diferencia es que una cuerda sólo puede estirar, mientras que una varilla puede estirar y empujar para garantizar que el péndulo se mueve sobre una circunferencia. Cuando las oscilaciones son pequeñas esto es irrelevante, pues la cuerda no hace sino estirar del péndulo en todo momento, pero, por ejemplo, si levantamos un péndulo tensando la cuerda con un ángulo mayor de 90° y lo dejamos caer, durante un tramo caerá verticalmente, hasta que la cuerda se vuelva a tensar y empiece a estirar, mientras que con una varilla la trayectoria será circular en todo momento.

Multiplicando la ecuación por θ' e integrando ambos miembros llegamos a

$$\frac{1}{2}\theta'^2 = \frac{g}{l} \cos \theta + C,$$

con lo que pasamos de una ecuación de segundo orden a otra de primer orden. Alternativamente, podemos usar el principio de conservación de la energía mecánica, lo cual nos da la interpretación física de la constante de integración C . En efecto, observemos que la fuerza que la varilla ejerce sobre la masa es tangente a su trayectoria, por lo que no produce trabajo y la energía mecánica de la partícula permanece constante. La velocidad es $v = l|\theta'|$, luego la energía cinética es $E_c = (1/2)ml^2\theta'^2$ y la energía mecánica es

$$E = \frac{1}{2}ml^2\theta'^2 - mgl \cos \theta$$

luego

$$\theta'^2 = \frac{2E}{ml^2} + \frac{2g}{l} \cos \theta, \quad (6.12)$$

que es la misma ecuación diferencial, pero ahora sabemos que la constante es $C = 2E/ml^2$. En realidad la dependencia de la masa es sólo aparente. Para verlo observamos que una forma conveniente de calcular E para un péndulo dado es considerar su velocidad v_0 cuando pasa por el punto más bajo, de modo que, en ese instante (luego siempre) $E = (1/2)mv_0^2 - mgl$, y así la ecuación es

$$\theta'^2 = \frac{v_0^2}{l^2} - \frac{2g}{l}(1 - \cos \theta) = \frac{v_0^2}{l^2} - \frac{4g}{l} \operatorname{sen}^2 \frac{\theta}{2}. \quad (6.13)$$

A partir de aquí distinguimos varios casos:

Si $E = -mgl$, tiene que ser $\theta' = 0$ en todo momento, luego el péndulo está en reposo. En particular, $v_0 = 0$ y la ecuación nos da que $\cos \theta = 1$, luego $\theta = 0$, es decir, que el péndulo se encuentra en su punto más bajo, como cabía esperar.

Supongamos ahora que $-mgl < E < mgl$. Entonces existe un único ángulo $0 < \theta_0 < \pi$ tal que

$$E = \frac{1}{2}ml^2\theta'^2 - mgl \cos \theta = -mgl \cos \theta_0. \quad (6.14)$$

Vemos entonces que el ángulo θ sólo puede variar en el intervalo $-\theta_0 \leq \theta \leq \theta_0$ y θ' sólo se anula cuando $\theta = \pm\theta_0$. Esto significa que el péndulo avanza hasta llegar a la altura θ_0 , momento en el que su velocidad llega a cero, y desde ahí empieza a caer hasta llegar a $-\theta_0$, y entonces vuelve a retroceder hasta llegar a θ_0 , y así sucesivamente, es decir, que el péndulo oscila.

En términos de v_0 , este caso se da cuando $-mgl < (1/2)mv_0^2 - mgl < mgl$, lo cual equivale a que $0 < v_0 < 2\sqrt{gl}$, y entonces

$$-mgl \cos \theta_0 = \frac{1}{2}mv_0^2 - mgl,$$

luego

$$\cos \theta_0 = 1 - \frac{v_0^2}{2gl}.$$

Equivalentemente

$$k = \operatorname{sen} \frac{\theta_0}{2} = \sqrt{\frac{1 - \cos \theta_0}{2}} = \frac{v_0}{2\sqrt{lg}}.$$

Sustituyendo $v_0^2 = 4lgk^2$ en (6.13) obtenemos:

$$\theta' = \pm 2\sqrt{\frac{g}{l}}(k^2 - \operatorname{sen} \frac{\theta}{2}).$$

El signo será positivo cuando el péndulo se desplaza en un sentido y negativo cuando lo hace en sentido opuesto. Vamos a estudiar, concretamente, el movimiento desde la posición $\theta = -\theta_0$ hasta $\theta = \theta_0$. Podemos medir el tiempo de modo que $\theta(0) = 0$, y así, si llamamos T al periodo de oscilación, la función $\theta(t)$ biyecta el semiperiodo $[-T/4, T/4]$ con el intervalo $[-\theta_0, \theta_0]$, por lo que podemos considerar la función inversa $t(\theta)$, cuya derivada es la inversa de la dada por la fórmula precedente:

$$\frac{dt}{d\theta} = \frac{1}{2} \sqrt{\frac{l}{g}} \frac{1}{\sqrt{k^2 - \operatorname{sen}^2(\theta/2)}}.$$

Separando las variables obtenemos que

$$t(\theta) = \int_0^\theta dt = \frac{1}{2} \sqrt{\frac{l}{g}} \int_0^\theta \frac{d\theta}{\sqrt{k^2 - \operatorname{sen}^2(\theta/2)}}, \quad (6.15)$$

donde hemos usado que $\theta = 0$ en $t = 0$. Para calcular la integral observamos que

$$-1 \leq \frac{\operatorname{sen}(\theta/2)}{\operatorname{sen}(\theta_0/2)} \leq 1,$$

por lo que podemos tomar el arco seno $-\pi/2 \leq \phi \leq \pi/2$ del cociente, y así cada ángulo θ se corresponde con un único ϕ que cumple

$$\operatorname{sen} \frac{\theta}{2} = k \operatorname{sen} \phi.$$

Despejando el seno de la derecha y derivando el arco seno obtenemos que

$$d\theta = \frac{2k \cos \phi}{\sqrt{1 - k^2 \operatorname{sen}^2 \phi}} d\phi$$

Por consiguiente, la integral (6.15) se puede transformar en

$$t(\theta) = \sqrt{\frac{l}{g}} \int_0^\phi \frac{k \cos \phi d\phi}{\sqrt{k^2 \cos^2 \phi \sqrt{1 - k^2 \operatorname{sen}^2 \phi}}} = \sqrt{\frac{l}{g}} \int_0^\phi \frac{d\phi}{\sqrt{1 - k^2 \operatorname{sen}^2 \phi}}.$$

En definitiva:

$$t(\theta) = \sqrt{l/g} F(\phi(\theta) | k^2), \quad (6.16)$$

donde $F(\phi | k)$ es la integral elíptica de primera especie. Equivalentemente, en términos de la amplitud de Jacobi:

$$\phi(\theta) = \operatorname{am}(\sqrt{g/l} t | k^2).$$

Por consiguiente,

$$\operatorname{sen} \frac{\theta}{2} = k \operatorname{sen} \phi = k \operatorname{sen} \operatorname{am}(\sqrt{g/l} t | k^2) = k \operatorname{sn}(\sqrt{g/l} t | k^2), \quad (6.17)$$

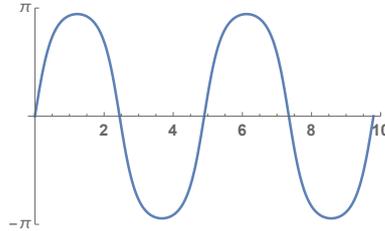
donde —recordemos— $k = \operatorname{sen}(\theta_0/2)$. La simetría de la función sn implica inmediatamente que esta fórmula es válida para todo tiempo t , y determina la función

$$\theta(t) = 2 \operatorname{arcsen}(k \operatorname{sn}(\sqrt{g/l} t | k^2)).$$

Claramente, el periodo de oscilación es

$$T = 4\sqrt{l/g} K(k^2).$$

La gráfica muestra la función $\theta(t)$ cuando se deja caer un péndulo con $l = 1$ desde una altura de $\theta_0 = 170^\circ$:



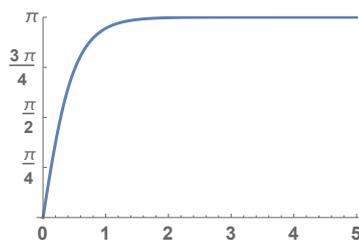
El periodo es $T = 4.896$ s.

Consideremos ahora el caso en el que la energía del péndulo es exactamente $E = mgl$. Una posibilidad entonces es que el péndulo se encuentre en reposo con la varilla en vertical, pero por encima del punto fijo ($\theta = \pi$). Esta solución es teóricamente posible, pero, en la práctica, la más mínima perturbación que modificara la posición de la varilla haría caer al péndulo, por lo que se dice que es un estado de *equilibrio inestable*.

La otra posibilidad es que, en un instante dado, el péndulo se encuentre en una posición distinta de $\theta = \pi$, con lo que $\theta' \neq 0$ (pues la energía potencial sería inferior a mgl). Este caso se da cuando $v_0 = 2\sqrt{gl}$ y puede tratarse exactamente igual que el anterior, sólo que ahora $\theta_0 = \pi$, $k = 1$ y, al llegar a la fórmula (6.16), nos encontramos con que la integral que define a $t(\theta)$ es divergente cuando $\theta \rightarrow \pi$, de modo que el péndulo tarda infinito tiempo en llegar a la posición vertical. Así pues, en teoría, un péndulo puede subir hasta quedarse parado en posición vertical. El movimiento sigue la ecuación correspondiente a (6.17) para $k = 1$, a saber:

$$\operatorname{sen} \frac{\theta}{2} = \tanh \sqrt{\frac{g}{l}} t.$$

La gráfica siguiente muestra la función $\theta(t)$ para un péndulo de 1 m de longitud con energía $E = mgl$ ($v_0 = 6.261$ m/s). Para $t = 4.4$ la diferencia de inclinación con la vertical es de 0.86 segundos de arco.



Finalmente, si $E > mgl$ entonces θ' nunca puede valer 0, luego el péndulo no oscila, sino que da vueltas completas. Vamos a analizar el movimiento en el intervalo en que θ avanza desde $-\pi$ hasta π en el intervalo de tiempo $[-T/2, T/2]$. En (6.13) tenemos que $\theta/2$ recorre el intervalo $[-\pi/2, \pi/2]$, luego $u = \operatorname{sen}(\theta/2)$ recorre el intervalo $[-1, 1]$. Observemos que $u' = \cos(\theta/2)\theta'/2$, luego $u'^2 = (1 - u^2)\theta'^2/4$ y al sustituir en (6.13) queda

$$\frac{u'^2}{(1 - u^2)} = \frac{v_0^2}{4l^2} - \frac{g}{l}u^2,$$

luego

$$u'^2 = \frac{v_0^2}{4l^2}(1 - k^2u^2)(1 - u^2),$$

donde ahora $k = 2\sqrt{lg}/v_0$. Como $u : [-T/2, T/2] \rightarrow [-1, 1]$ es biyectiva, podemos considerar la función inversa $t(u)$, cuya derivada es

$$\frac{dt}{du} = \frac{2l}{v_0} \frac{1}{\sqrt{(1 - k^2u^2)(1 - u^2)}}.$$

Ahora separamos variables e integramos usando que $t(0) = 0$:

$$\frac{v_0}{2l}t = \frac{v_0}{2l} \int_0^t dt = \int_0^u \frac{du}{\sqrt{(1-k^2u^2)(1-u^2)}}.$$

Pero por (6.7) esto es equivalente a

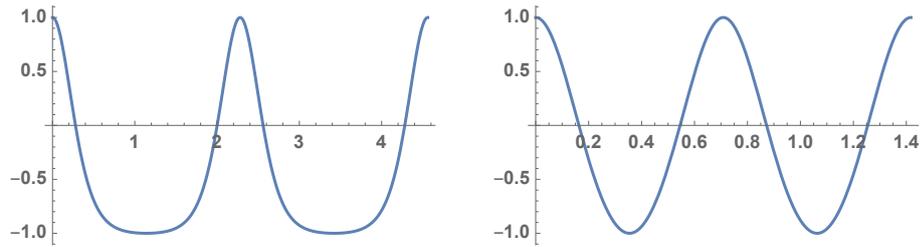
$$\operatorname{sen} \frac{\theta}{2} = u = \operatorname{sn}(v_0 t / 2l | k^2),$$

o, eliminando los senos,

$$\theta = 2 \operatorname{am}(v_0 t / 2l | k^2).$$

Vemos que θ avanza 2π cuando la amplitud avanza π , lo cual sucede cuando su argumento avanza $2K(k^2)$, lo que a su vez se corresponde con un lapso de tiempo $T = (4l/v_0)K(k^2)$.

Las gráficas muestran la función $\theta(t)$ para un péndulo con $l = 1$ y velocidades $v_0 = 6.3 \text{ m/s}$ (ligeramente superior a la velocidad mínima de rotación, 6.261 m/2) y $v_0 = 10 \text{ m/2}$, respectivamente.



Vemos que en el primer caso la velocidad es muy pequeña durante un lapso de tiempo cercano a la mitad del periodo de rotación, $T = 2.28 \text{ s}$. El periodo en el segundo caso es $T = 0.71 \text{ s}$.

La tabla siguiente resume todos los casos:

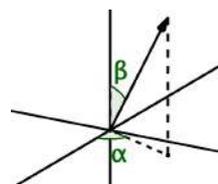
E	v_0	k	$\theta(t)$	T
$-mgl$	0	0	0	0
$] -mgl, mgl[$	$] 0, 2\sqrt{lg}[$	$\frac{v_0}{2\sqrt{lg}}$	$2 \operatorname{arcsen}(k \operatorname{sn}(\sqrt{\frac{g}{l}} t k^2))$	$4\sqrt{\frac{l}{g}} K(k^2)$
mgl	$2\sqrt{lg}$	1	$2 \operatorname{arcsen} \tanh \sqrt{\frac{g}{l}} t$	∞
$> mgl$	$> 2\sqrt{lg}$	$\frac{2\sqrt{lg}}{v_0}$	$2 \operatorname{am}(\frac{v_0 t}{2l} k^2)$	$\frac{4l}{v_0} K(k^2)$

Además, en el segundo caso la amplitud de la oscilación viene dada por $\operatorname{sen}(\theta_0/2) = k$. La tabla no incluye el caso del equilibrio inestable, correspondiente a $E = mgl$ con $\theta(t) = \pi$.

6.6 El péndulo esférico

Ahora vamos a estudiar el movimiento de un péndulo sin imponer que su trayectoria deba estar contenida en un plano. En cualquier caso, estará contenida en la esfera de centro el punto fijo O y radio l , por lo que un péndulo con movimiento no plano se conoce como un *péndulo esférico*.

Para estudiarlo consideraremos *coordenadas esféricas*, de modo que β será el ángulo que la varilla forma con el semi-eje vertical positivo y α el ángulo que su proyección en el plano XY forma con el semi-eje X positivo. Así, el vector de posición de la partícula es



$$\vec{r} = (l \cos \alpha \sin \beta, l \sin \alpha \sin \beta, l \cos \beta), \quad (6.18)$$

y la velocidad es

$$\vec{v} = l(\beta' \cos \alpha \cos \beta - \alpha' \sin \alpha \sin \beta, \alpha' \cos \alpha \sin \beta + \beta' \sin \alpha \cos \beta, -\beta' \sin \beta).$$

A partir de aquí podemos calcular el momento angular $\vec{L} = \vec{r} \times m\vec{v}$, pero sólo necesitamos su tercera componente:

$$L_3 = I\alpha' \sin^2 \beta, \quad (6.19)$$

donde hemos llamado³ $I = ml^2$. En general,

$$\frac{d\vec{L}}{dt} = \vec{v} \times m\vec{v} + \vec{r} \times m\vec{a} = \vec{0} + \vec{r} \times \vec{F},$$

donde \vec{F} es la fuerza total que actúa sobre el péndulo, que es la suma del peso y de la varilla de la cuerda. El momento total $\vec{M} = \vec{r} \times \vec{F}$ será la suma de los momentos de ambas fuerzas, pero el momento de la tensión es nulo porque tiene la dirección de \vec{r} , así que \vec{M} coincide con el momento del peso,

$$\vec{M} = \vec{r} \times \vec{P} = -mgl \sin \beta (\sin \alpha, -\cos \alpha, 0).$$

El hecho de que $M_3 = 0$ se traduce en que L_3 es constante, lo que convierte a (6.19) en una ecuación diferencial que relaciona a α' y a β . Obtendremos otra a partir del principio de conservación de la energía, pero antes observemos que podemos descartar el caso $L_3 = 0$, ya que corresponde al péndulo simple que hemos estudiado en la sección anterior.

En efecto, un caso en el que $L_3 = 0$ se da si $\sin \beta$ es constante igual a 0, lo cual significa que el péndulo está en equilibrio estable o inestable, luego es un caso trivial de péndulo simple. Si no se da este caso, vamos a probar que α' es constante igual a 0, lo cual equivale a que α es constante, y esto se interpreta como que el péndulo se mueve en un plano vertical, luego es un péndulo simple.

³No necesitamos este hecho, pero I es el momento de inercia del péndulo.

En principio, lo que sabemos es que $\alpha' = 0$ siempre que $\sin \beta \neq 0$, y sólo falta probar que lo mismo sucede cuando $\sin \beta = 0$. En efecto, si t_0 es un instante en el que $\sin \beta = 0$, para todo $n > 0$, no puede ser que $\sin \beta = 0$ en todo el intervalo $]t_0 - 1/n, t_0 + 1/n[$, pues eso significaría que el péndulo está en equilibrio durante un intervalo de tiempo, pero si está en equilibrio durante un intervalo de tiempo, lo está en todo momento. Por lo tanto, existe un instante t_n tal que $|t_n - t_0| < 1/n$ y $\sin \beta(t_n) \neq 0$, luego $\alpha'(t_n) = 0$. La sucesión $\{t_n\}_{n=1}^{\infty}$ tiende a t_0 , luego por continuidad $\alpha'(t_n) = 0$ tiende a $\alpha'(t_0) = 0$.

Así pues, de aquí en adelante suponemos que $L_3 \neq 0$. En particular, (6.19) implica que $\sin \beta \neq 0$ en todo momento.

Un simple cálculo muestra que la energía mecánica del péndulo es

$$E = \frac{1}{2}mv^2 + mgl \cos \beta = \frac{1}{2}I(\beta'^2 + \alpha'^2 \sin^2 \beta) + mgl \cos \beta.$$

Despejando α' en (6.19) y sustituyendo en la ecuación anterior obtenemos

$$E = \frac{1}{2}I\beta'^2 + \frac{L_3^2}{2I \sin^2 \beta} + mgl \cos \beta.$$

Llamando $u = \cos \beta$, tenemos que $\beta'^2 = u'^2/(1 - u^2)$, con lo que obtenemos la ecuación diferencial

$$u'^2 = f(u), \quad (6.20)$$

donde

$$f(u) = \frac{2}{I}(E - mglu)(1 - u^2) - \frac{L_3^2}{I^2}. \quad (6.21)$$

El péndulo cónico Vamos a estudiar primeramente un caso particular de movimiento del péndulo esférico, a saber, el caso en que $\beta' = 0$, de modo que éste se encuentra siempre a la misma altura.

La ecuación (6.19) muestra que α' ha de ser constante, con lo que el péndulo describirá una trayectoria circular horizontal con velocidad angular constante. Para calcular esta velocidad angular α' observamos que la velocidad del péndulo se reduce a

$$\vec{v} = l(-\alpha' \sin \alpha \sin \beta, \alpha' \cos \alpha \sin \beta, 0),$$

y entonces es fácil calcular el momento angular:

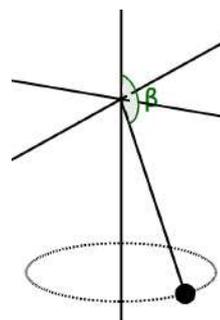
$$\vec{L} = I\alpha' \sin \beta(-\cos \alpha \cos \beta, -\sin \alpha \cos \beta, \sin \beta).$$

Su derivada es

$$\frac{d\vec{L}}{dt} = I\alpha'^2 \sin \beta \cos \beta (\sin \alpha, -\cos \alpha, 0)$$

y al igualar a \vec{M} obtenemos las ecuaciones

$$l\alpha'^2 \cos \beta \sin \alpha = -g \sin \alpha, \quad l\alpha'^2 \cos \beta \cos \alpha = -g \cos \alpha.$$



De aquí se sigue la relación

$$\alpha'^2 = -\frac{g}{l \cos \beta}, \quad (6.22)$$

que determina la velocidad angular necesaria para que el ángulo β pueda mantenerse constante. En particular vemos que ha de ser $\cos \beta < 0$, es decir, que $\beta > \pi/2$, lo que significa que el plano de giro ha de ser inferior al punto fijo O , como era de prever. Además, cuando más elevado esté el plano de giro, mayor ha de ser la velocidad angular. El tiempo que tarda el péndulo en dar una vuelta completa es

$$T = 2\pi \sqrt{\frac{l}{g}} \sqrt{-\cos \beta}.$$

En principio, la condición $\beta' = 0$ que estamos imponiendo equivale a que sea $u' = 0$ y, en virtud de (6.20), a que $u = \cos \beta$ sea raíz del polinomio $f(u)$. Terminaremos el análisis del péndulo cónico demostrando que $\beta' = 0$ implica de hecho que $u = \cos \beta$ es raíz doble de $f(u)$. En efecto, la condición $f(u) = 0$ equivale a que

$$2(E - mglu)(1 - u^2) = \frac{L_3^2}{I}.$$

Bajo esta hipótesis, u será una raíz doble⁴ si y sólo si

$$f'(u) = -2mgl(1 - u^2) - 4(E - mglu)u = 0.$$

Despejando $E - mglu$ en la primera ecuación y sustituyendo en la segunda vemos que ésta equivale a que

$$-2mgl(1 - u^2) = \frac{2L_3^2 u}{I(1 - u^2)}.$$

Ahora bien, por definición de L_3 tenemos que $L_3^2 = I^2 \alpha'^2 (1 - u^2)^2$, luego esta ecuación equivale a que

$$-mgl = I \alpha'^2 u,$$

que a su vez se simplifica hasta (6.22), luego, en efecto, $u = \cos \beta$ es raíz doble. ■

El caso general Las características cualitativas del movimiento del péndulo esférico se deducen fácilmente del análisis del polinomio $f(u)$ dado por (6.21), que, alternativamente, podemos expresar como

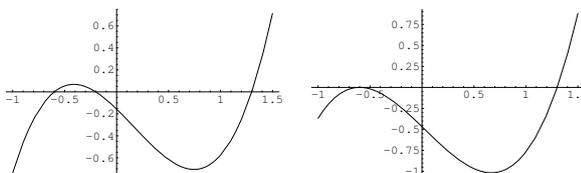
$$f(u) = \left(\frac{2E}{I} - \frac{2g}{l}u\right)(1 - u^2) - \frac{L_3^2}{I^2}.$$

⁴En general, si c es una raíz de un polinomio $f(x)$, entonces $f(x) = (x - c)^n g(x)$, donde $g(c) \neq 0$, y $f'(x) = n(x - c)^{n-1} g(x) + (x - c)^n g'(x)$, luego la raíz es doble si y sólo si $n \geq 2$, si y sólo si $f'(c) = 0$.

Se trata de un polinomio cúbico con coeficiente director $2g/l > 0$, por lo que tiende a $\pm\infty$ cuando u tiende a $\pm\infty$ respectivamente. Además, recordando que estamos suponiendo $L_3 \neq 0$,

$$f(\pm 1) = -\frac{L_3^2}{l^2} < 0$$

y como, según (6.20), para $u = \cos \beta$ ha de ser $f(u) = u'^2 \geq 0$, el polinomio ha de ser no negativo al menos en un punto del intervalo $[-1, 1]$. Estas condiciones hacen que la gráfica de $f(u)$ tenga que tener el aspecto de una de las dos figuras siguientes:



Llamaremos $u_1 \leq u_2 \leq u_3$ a las raíces de $f(u)$, entendiendo que dos de ellas son iguales si se trata de una raíz doble. Más precisamente, las condiciones que hemos indicado implican que

$$-1 < u_1 \leq u_2 < 1 < u_3.$$

La segunda gráfica corresponde al caso $u_1 = u_2$. En ese caso, necesariamente $\cos \beta$ toma el valor constante u_1 , luego β es constante y se trata del caso del péndulo cónico que ya hemos estudiado. Podemos suponer, pues, que

$$-1 < u_1 < u_2 < 1 < u_3.$$

Entonces $u = \cos \beta$ ha de oscilar periódicamente entre u_1 y u_2 . En efecto, si, por ejemplo, en un instante dado se cumple que $u' > 0$, entonces u deberá crecer hasta que $u'^2 = f(u)$ se anule, lo cual sucede cuando llega a u_2 y, como no puede seguir adelante ni mantenerse constante (porque entonces estaríamos en el caso del péndulo cónico, lo cual no es posible porque u_2 no es una raíz doble de $f(u)$), necesariamente ha de pasar a decrecer (es decir, u' ha de pasar a ser negativa) hasta llegar a u_1 , momento en el cual debe pasar a crecer, y así sucesivamente.

Factorizando

$$f(u) = \frac{2g}{l}(u - u_1)(u_2 - u)(u_3 - u)$$

y comparando los coeficientes, vemos que

$$u_1 + u_2 + u_3 = \frac{E}{mgl}, \quad u_1u_2 + u_1u_3 + u_2u_3 = -1.$$

La segunda ecuación nos permite despejar

$$u_3 = -\frac{1 + u_1u_2}{u_1 + u_2},$$

que nos determina la tercera raíz de $f(u)$ a partir de las raíces correspondientes a los extremos de la oscilación.

Pongamos que el péndulo va u_1 a u_2 en el intervalo de tiempo $[0, T/2]$. En dicho intervalo (salvo en los extremos) $u' > 0$, luego $u(t)$ es creciente y podemos considerar su inversa $t(u)$, que, teniendo en cuenta (6.20), cumplirá la ecuación diferencial

$$\frac{dt}{du} = \frac{1}{\sqrt{f(u)}}.$$

Separando las variables tenemos que

$$t = \int_{u_1}^u \frac{du}{\sqrt{f(u)}} = \sqrt{\frac{l}{2g}} \int_{u_1}^u \frac{du}{\sqrt{(u-u_1)(u_2-u)(u_3-u)}}. \quad (6.23)$$

Hacemos el cambio de variable $z = \sqrt{u-u_1}$ (con lo que z varía en el intervalo $[0, \sqrt{u_2-u_1}]$):

$$t = \sqrt{\frac{l}{2g}} \int_0^z \frac{2z dz}{z\sqrt{(u_2-u_1-z^2)(u_3-u_1-z^2)}}.$$

Ahora hacemos el cambio $z = \sqrt{u_2-u_1} w$ (con lo que w varía en $[0, 1]$):

$$\begin{aligned} t &= \sqrt{\frac{2l}{g}} \int_0^w \frac{\sqrt{u_2-u_1} dw}{\sqrt{((u_2-u_1)-(u_2-u_1)w^2)((u_3-u_1)-(u_2-u_1)w^2)}} \\ &= \sqrt{\frac{2l}{g(u_3-u_1)}} \int_0^w \frac{dw}{\sqrt{(1-w^2)(1-\frac{u_2-u_1}{u_3-u_1}w^2)}} = \frac{1}{p} \int_0^w \frac{dw}{\sqrt{(1-w^2)(1-k^2w^2)}}, \end{aligned}$$

donde

$$p = \sqrt{\frac{g(u_3-u_1)}{2l}}, \quad k = \sqrt{\frac{u_2-u_1}{u_3-u_1}}.$$

Ahora (6.7) implica que $w = \text{sn}(pt | k)$, luego

$$z = \sqrt{u_2-u_1} \text{sn}(pt | k^2)$$

y, finalmente,

$$\cos \beta = u = u_1 + (u_2-u_1) \text{sn}^2(pt | k^2). \quad (6.24)$$

Esta ecuación determina la función $\beta(t)$. La función sn tiene periodo $4K(k^2)$, pero sn^2 tiene periodo $2K$. Por lo tanto, el periodo de la oscilación es

$$T = \frac{2K(k^2)}{p} = \sqrt{\frac{8l}{g(u_3-u_1)}} K(k^2).$$

Ahora falta estudiar la función $\alpha(t)$. En principio, por (6.19) tenemos que

$$\alpha' = \frac{L_3}{I(1-u^2)}, \quad (6.25)$$

luego, tomando el sistema de referencia de modo que $\alpha(0) = 0$, la función $\alpha(t)$ viene dada por

$$\alpha(t) = \int_0^t \frac{L_3 dt}{I(1-u^2)} = \frac{L_3}{I} \int_{u_1}^u \frac{du}{(1-u^2)\sqrt{f(u)}}.$$

En términos de las raíces de $f(u)$ la expresión es

$$\alpha(u) = \frac{L_3}{I} \sqrt{\frac{l}{2g}} \int_{u_1}^u \frac{du}{(1-u^2)\sqrt{(u-u_1)(u_2-u)(u_3-u)}},$$

y como $L_3/I = \sqrt{-f(1)} = \sqrt{2g/l} \sqrt{(1-u_1)(1-u_2)(u_3-1)}$,

$$\alpha(u) = \int_{u_1}^u \frac{\sqrt{(1-u_1)(1-u_2)(u_3-1)}}{(1-u^2)\sqrt{(u-u_1)(u_2-u)(u_3-u)}} du. \quad (6.26)$$

Esta integral no puede expresarse ni siquiera en términos de integrales elípticas, pero, como cualquier otra integral, puede aproximarse numéricamente.

Ejemplo Vamos a estudiar el movimiento de un péndulo esférico con una varilla de $l = 1$ m de longitud que alcance su punto más alto (es decir, con $\beta' = 0$) en $\beta_0 = 60^\circ$ y con una velocidad $v_0 = 5$ m/s. En realidad la masa del péndulo es irrelevante, pues se cancela en todas las fórmulas finales, pero, por concretar los valores de la energía, etc., podemos suponer que $m = 1$ kg.

Entonces $I = 1 \text{ kg}\cdot\text{m}^2$, $E = \frac{1}{2}mv_0^2 + mg \cos \beta_0 = 17.4$ J, la relación

$$v^2 = l^2(\beta'^2 + \alpha'^2 \sin^2 \beta)$$

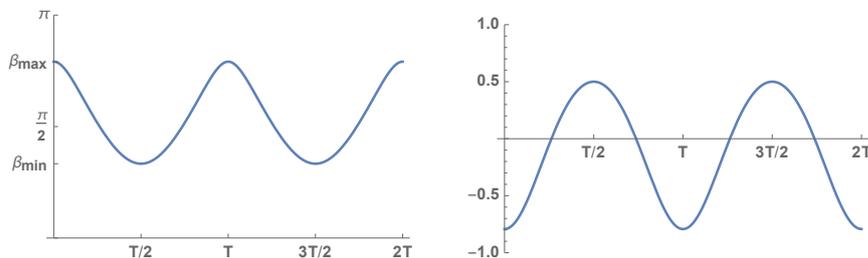
nos da que $\alpha'_0 = 10/\sqrt{3}$, de donde $L_3 = I\alpha' \sin^2 \beta = 5\sqrt{3}/2$. Por lo tanto:

$$f(u) = 2(1-u^2)(17.4 - 9.8u) - \frac{75}{4},$$

cuyas raíces son $u_1 = -0.7921$, $u_2 = \cos \beta_0 = 1/2$, $u_3 = 2.06761$. De aquí se desprende que la altitud mínima del péndulo será $\arccos u_1 = 142.38^\circ$, es decir, 37.81° respecto al semieje vertical negativo. Además, de aquí deducimos las constantes

$$k^2 = 0.4518, \quad p = 3.7433, \quad T = 0.9699 \text{ s}.$$

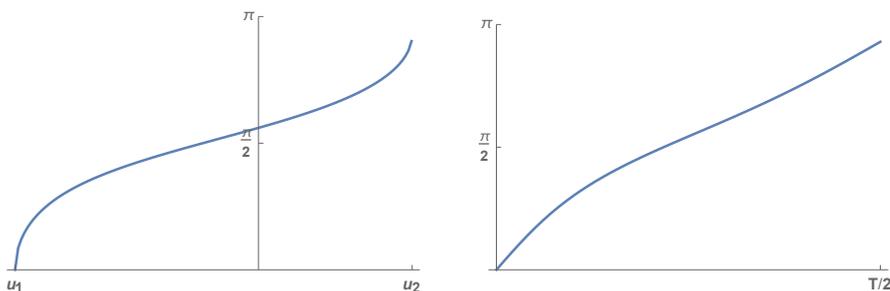
Así (6.24) nos da las funciones $\beta(t)$ y $u(t) = \cos \beta(t)$:



Notemos que β_{\max} corresponde a la altitud mínima del péndulo, pues β se mide respecto del semieje vertical positivo y cuanto mayor es β más abajo está el péndulo. Ahora:

$$\alpha(u) = \int_{u_1}^u \frac{0.978076}{(1-u^2)\sqrt{(u+0.7921)(0.5-u)(2.06761-u)}} du.$$

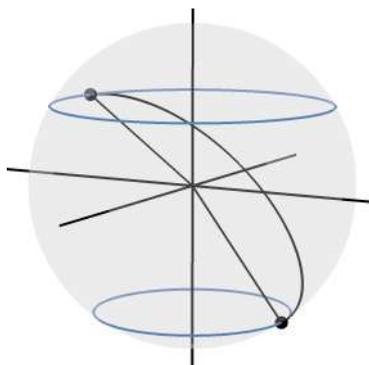
La gráfica de la izquierda se ha obtenido calculando numéricamente $\alpha(u)$ para valores de u entre u_1 y u_2 de centésima en centésima, mientras que la de la derecha es $\alpha(t) = \alpha(u(t))$:



En particular,

$$\Delta\alpha = \int_{u_1}^{u_2} \frac{0.978076}{(1-u^2)\sqrt{(u+0.7921)(0.5-u)(2.06761-u)}} du$$

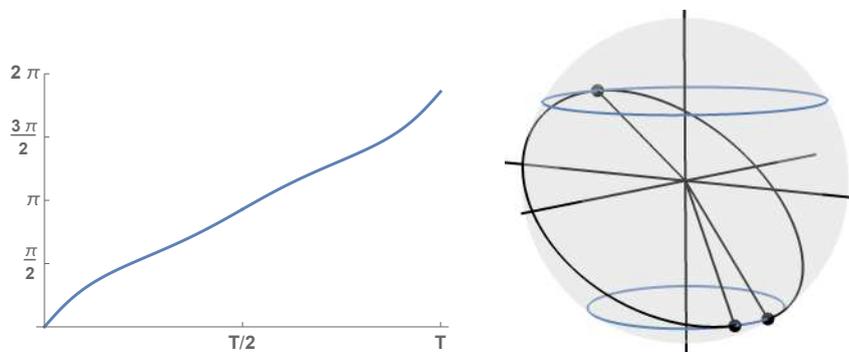
resulta ser $\Delta\alpha = 167.3^\circ$, lo que significa que, mientras el péndulo sube desde su posición más baja en $t = 0$ hasta su posición más alta en $t = T/2$, gira 167.3° respecto al eje vertical, es decir, no llega dar una vuelta entera. La figura muestra el arco $r(t)$ determinado por (6.18):



En el intervalo de tiempo $[T/2, T]$, el valor de α vendrá dado por

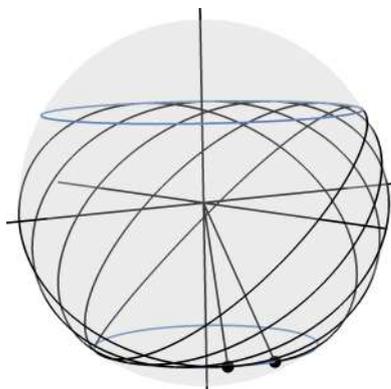
$$\Delta\alpha + \int_u^{u_2} \frac{0.978076}{(1-u^2)\sqrt{(u+0.7921)(0.5-u)(2.06761-u)}} du,$$

o, equivalentemente, $2\Delta\alpha - \alpha(u)$. Esto nos determina la función $\alpha(t)$ en todo el intervalo $[0, T]$, que es la que muestra la gráfica de la izquierda. A la derecha vemos el recorrido del péndulo en este intervalo:



En el instante $t = T$, el péndulo se encuentra exactamente en la misma situación (de posición y velocidad) que en $t = 0$ salvo que su posición se ha desplazado un ángulo de $360^\circ - 2\Delta\alpha = 25.4^\circ$, es decir, que el punto de partida no es ahora el mismo que en $t = 0$, sino que ha sufrido una *precesión*.

A partir de ahí, el péndulo describirá un arco idéntico al que hemos calculado, salvo que estará desplazado 25.5° , y así sucesivamente. La figura siguiente muestra la trayectoria hasta $t = 5T$:



Ahora vamos a probar que el comportamiento que hemos obtenido en el ejemplo anterior no es casual, sino que siempre se produce una precesión, es decir, vamos a demostrar que siempre se cumple que

$$\frac{\pi}{2} < \Delta\alpha < \pi,$$

con lo que la precesión es siempre $0 < 2\pi - 2\Delta\alpha < \pi$. En principio:

$$\Delta\alpha = \int_{u_1}^{u_2} \frac{\sqrt{(1-u_1)(1-u_2)(u_3-1)}}{(1-u^2)\sqrt{(u-u_1)(u_2-u)(u_3-u)}} du. \quad (6.27)$$

Como $1 - u^2 = (1 + u)(1 - u) > (1 - u)$ y

$$\frac{1}{\sqrt{u_3 - u}} < \frac{1}{\sqrt{u_3 - 1}},$$

tenemos que

$$\begin{aligned} \Delta\alpha &< \int_{u_1}^{u_2} \frac{\sqrt{(1 - u_1)(1 - u_2)(u_3 - 1)}}{(1 - u)\sqrt{(u - u_1)(u_2 - u)(u_3 - u)}} du \\ &< \int_{u_1}^{u_2} \frac{\sqrt{(1 - u_1)(1 - u_2)}}{(1 - u)\sqrt{(u - u_1)(u_2 - u)}} du. \end{aligned}$$

Demostraremos que la última integral es π y esto nos dará una de las desigualdades que buscamos. Por otro lado, descomponemos

$$\frac{1}{1 - u^2} = \frac{1}{2} \frac{1}{1 + u} + \frac{1}{2} \frac{1}{1 - u},$$

con lo que la integral se descompone a su vez en

$$\begin{aligned} \Delta\alpha &= \frac{1}{2} \int_{u_1}^{u_2} \frac{\sqrt{(1 - u_1)(1 - u_2)(u_3 - 1)}}{(1 + u)\sqrt{(u - u_1)(u_2 - u)(u_3 - u)}} du \\ &+ \frac{1}{2} \int_{u_1}^{u_2} \frac{\sqrt{(1 - u_1)(1 - u_2)(u_3 - 1)}}{(1 - u)\sqrt{(u - u_1)(u_2 - u)(u_3 - u)}} du. \end{aligned}$$

Es evidente que la segunda integral es positiva, porque el integrando lo es. Además, como $u > -1$, se cumple

$$\frac{1}{\sqrt{u_3 - u}} > \frac{1}{\sqrt{u_3 + 1}},$$

con lo que

$$\begin{aligned} \Delta\alpha &> \frac{1}{2\sqrt{u_3 + 1}} \int_{u_1}^{u_2} \frac{\sqrt{(1 - u_1)(1 - u_2)(u_3 - 1)}}{(1 + u)\sqrt{(u - u_1)(u_2 - u)}} du \\ &= \frac{1}{2\sqrt{u_3 + 1}} \int_{u_1}^{u_2} \frac{\sqrt{(u_1 + 1)(u_2 + 1)(u_3 + 1)}}{(1 + u)\sqrt{(u - u_1)(u_2 - u)}} du \\ &= \frac{1}{2} \int_{u_1}^{u_2} \frac{\sqrt{(u_1 + 1)(u_2 + 1)}}{(1 + u)\sqrt{(u - u_1)(u_2 - u)}} du, \end{aligned}$$

donde hemos usado que $f(1) = f(-1)$ para cambiar el numerador del integrando.

De este modo, basta probar que

$$\int_{u_1}^{u_2} \frac{\sqrt{(1 - u_1)(1 - u_2)} du}{(1 - u)\sqrt{(u - u_1)(u_2 - u)}} = \int_{u_1}^{u_2} \frac{\sqrt{(u_1 + 1)(u_2 + 1)} du}{(1 + u)\sqrt{(u - u_1)(u_2 - u)}} = \pi. \quad (6.28)$$

Las dos integrales se resuelven mediante las técnicas usuales de cálculo de primitivas, concretamente con el cambio de Euler

$$\sqrt{(u - u_1)(u_2 - u)} = t(u_2 - u).$$

Explícitamente:

$$u = \frac{u_1 + u_2 t^2}{1 + t^2}, \quad du = \frac{2t(u_2 - u_1)}{(1 + t^2)^2} dt.$$

En particular, la derivada de u muestra que la función $u(t)$ es creciente en $[0, +\infty[$, y biyecta este intervalo con $[u_1, u_2[$. Además:

$$\sqrt{(u - u_1)(u_2 - u)} = t \frac{u_2 - u_1}{1 + t^2}, \quad 1 \pm u = \frac{1 \pm u_1 + t^2(1 \pm u_2)}{1 + t^2}.$$

Así, por ejemplo, la primera integral se transforma en

$$\begin{aligned} & 2\sqrt{(1 + u_1)(1 + u_2)} \int_0^{+\infty} \frac{dt}{1 + u_1 + t^2(1 + u_2)} \\ &= 2 \int_0^{+\infty} \frac{\sqrt{\frac{1+u_2}{1+u_1}}}{1 + \left(\sqrt{\frac{1+u_2}{1+u_1}} t\right)^2} dt = \lim_{x \rightarrow +\infty} 2 \arctan \sqrt{\frac{1 + u_2}{1 + u_1}} x = \pi. \end{aligned}$$

Igualmente se razona con la segunda integral. ■

6.7 La media aritmético-geométrica

En [ITAn] (ejemplo 1.18) vimos el método de Herón para calcular raíces cuadradas, que muestra que si partimos de dos números reales $0 < b < a$ y vamos calculando sucesivamente la media harmónica y la media aritmética:

$$h_0 = b, \quad a_0 = a, \quad h_{n+1} = \frac{2}{\frac{1}{h_n} + \frac{1}{a_n}}, \quad a_{n+1} = \frac{h_n + a_n}{2},$$

obtenemos dos sucesiones monótonas que convergen a la media geométrica \sqrt{ab} de los números de partida. Gauss y Legendre estudiaron independientemente lo que sucede si, en lugar de calcular medias harmónicas y aritméticas, calculamos medias aritméticas y geométricas:

Dados dos números reales $0 < b \leq a$, definimos sucesiones

$$a_0 = a, \quad g_0 = b, \quad a_{n+1} = \frac{a_n + g_n}{2}, \quad g_{n+1} = \sqrt{a_n g_n}.$$

Por la desigualdad entre las medias aritmética y geométrica, $g_{n+1} \leq a_{n+1}$. Más aún, tenemos que

$$g_0 \leq g_1 \leq g_2 \leq \dots \leq a_2 \leq a_1 \leq a_0.$$

En efecto, como $g_n \leq a_n$, también $g_n^2 \leq a_n g_n$, luego

$$g_n \leq \sqrt{a_n g_n} = g_{n+1} \leq a_{n+1} = \frac{a_n + g_n}{2} \leq a_n.$$

Además, se cumple que

$$a_n - g_n \leq 2^{-n}(a - b), \tag{6.29}$$

pues

$$a_{n+1} - g_{n+1} \leq a_{n+1} - g_n = \frac{a_n + g_n}{2} - g_n = 2^{-1}(a_n - g_n),$$

y basta aplicar n veces esta desigualdad.

Esto implica que las dos sucesiones $\{a_n\}_{n=0}^\infty$ y $\{g_n\}_{n=0}^\infty$, no sólo convergen por ser monótonas y estar acotadas [ITAn 1.16], sino que ambas convergen al mismo límite, que representaremos por $M(a, b)$ y que recibe el nombre de *media aritmético-geométrica* de a y b .

La convergencia es muy rápida, de hecho, mucho más rápida de lo que se desprende de (6.29). En efecto, observemos que

$$a_{n+1}^2 - g_{n+1}^2 = \frac{a_n^2 + g_n^2 + 2a_n g_n}{4} - a_n g_n = \frac{(a_n - g_n)^2}{4},$$

luego

$$a_{n+1} - g_{n+1} = \frac{(a_n - g_n)^2}{4(a_{n+1} + g_{n+1})} = \frac{(a_n - g_n)^2}{8a_{n+2}} \leq \frac{(a_n - g_n)^2}{8b}.$$

Así, si $8b \geq 1$, tenemos que $a_{n+1} - g_{n+1} \leq (a_n - g_n)^2$, por lo que el número de cifras exactas se duplica en cada paso. En general podemos tomar un N tal que $a_N - g_N < 4b$, y así, si $n \geq N$,

$$\begin{aligned} |a_n - g_n| &\leq \frac{(a_{n-1} - g_{n-1})^2}{8b} \leq \frac{(a_{n-2} - g_{n-2})^4}{(8b)^{1+2}} \leq \frac{(a_{n-3} - g_{n-3})^8}{(8b)^{1+2+2^2}} \leq \dots \\ &\dots \leq \frac{(a_N - g_N)^{2^{n-N}}}{(8b)^{1+2+\dots+2^{n-N-1}}} = \frac{(a_N - g_N)^{2^{n-N}}}{(8b)^{2^{n-N}-1}} \leq \frac{4b}{2^{2^{n-N}-1}} = \frac{8b}{2^{2^{n-N}}}. \end{aligned}$$

En otros términos:

$$a_n - g_n = O(2^{-2^{n-N}}). \tag{6.30}$$

Por ejemplo, para calcular $M(\sqrt{2}, 1)$ construimos la tabla siguiente:

n	a_n	g_n
0	1.414213562373095048802	1.000000000000000000000
1	1.207106781186547524401	1.189207115002721066717
2	1.198156948094634295559	1.198123521493120122607
3	1.198140234793877209083	1.198140234677307205798
4	1.198140234735592207441	1.198140234735592207439

de donde deducimos que

$$M(\sqrt{2}, 1) = 1.1981402347355922074\dots$$

con todos los decimales exactos. Vemos que, como hemos anticipado, el número de cifras exactas en cada paso es más del doble de las del paso anterior.

La tabla la ha calculado un ordenador, pero aparece exactamente igual en un trabajo de Gauss. De hecho, Gauss no sólo calculó este valor, sino que se sorprendió porque reconoció el resultado, y el 30 de mayo de 1799 anotó en su diario:

He comprobado que la media aritmético-geométrica entre 1 y $\sqrt{2}$ es π/ϖ hasta la undécima cifra decimal. La demostración de este hecho abrirá sin duda una rama del análisis completamente nueva.

Antes de fin de año Gauss tenía dos demostraciones de un teorema general que implica en particular el caso concreto que había observado. Para probarlo consideramos la integral

$$I(a, b) = \int_0^{\pi/2} \frac{1}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}} d\phi = \frac{1}{a} K(m), \quad (6.31)$$

donde $m = 1 - (b/a)^2$. En efecto,

$$I(a, b) = \int_0^{\pi/2} \frac{1}{\sqrt{a^2 - (a^2 - b^2) \sin^2 \phi}} d\phi = \frac{1}{a} \int_0^{\pi/2} \frac{1}{\sqrt{1 - m \sin^2 \phi}} d\phi = \frac{K(m)}{a}.$$

En particular,

$$I(\sqrt{2}, 1) = \frac{1}{\sqrt{2}} K(1/2) = \int_0^1 \frac{1}{\sqrt{1-x^4}} dx = \frac{\varpi}{2}.$$

Por lo tanto, lo que Gauss había constatado era que

$$M(\sqrt{2}, 1)I(\sqrt{2}, 1) = \frac{\pi}{2},$$

y lo que demostró en general fue:

Teorema 6.17 *Si $a \geq b > 0$, entonces*

$$M(a, b) \int_0^{\pi/2} \frac{1}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}} d\phi = \frac{\pi}{2}.$$

DEMOSTRACIÓN: Llamemos $M = M(a, b)$ y vamos a probar que

$$\lim_n I(a_n, g_n) = \frac{\pi}{2M}.$$

Para ello observamos que

$$b^2 = g_0^2 \leq g_n^2 \leq a_n^2 \cos^2 \phi + g_n^2 \sin^2 \phi \leq a_n^2 \leq a_0^2 = a^2.$$

La función $f : [b^2, a^2] \rightarrow [1/a, 1/b]$ dada por $f(t) = 1/\sqrt{t}$ es continua, luego por el teorema de Heine-Cantor [ITAn A.14], dado $\epsilon > 0$, existe un $\delta > 0$ tal que si $t, t' \in [b^2, a^2]$ cumplen $|t - t'| < \delta$, entonces $|f(t) - f(t')| < \epsilon$.

Por otra parte, si $0 \leq \phi \leq \pi/2$,

$$\begin{aligned} |a_n^2 \cos^2 \phi + g_n^2 \sin^2 \phi - M^2| &= |a_n^2 \cos^2 \phi + g_n^2 \sin^2 \phi - M^2(\cos^2 \phi + \sin^2 \phi)| \\ &\leq |a_n^2 - M^2| + |g_n^2 - M^2|, \end{aligned}$$

y $\lim_n a_n = \lim_n b_n = M$, luego existe un n_0 tal que si $n \geq n_0$ entonces

$$|a_n^2 \cos^2 \phi + g_n^2 \sin^2 \phi - M^2| < \delta,$$

luego, aplicando f , también $|F_n(\phi) - M^{-1}| < \epsilon$. Por lo tanto,

$$\left| \int_0^{\pi/2} F_n(\phi) d\phi - \int_0^{\pi/2} \frac{1}{M} d\phi \right| \leq \int_0^{\pi/2} |F_n(\phi) - M^{-1}| d\phi \leq \int_0^{\pi/2} \epsilon d\phi = \frac{\pi\epsilon}{2}.$$

Esto prueba que

$$\lim_n I(a_n, g_n) = \int_0^{\pi/2} \frac{1}{M} d\phi = \frac{\pi}{2M}.$$

Para completar la prueba del teorema basta ver que la sucesión $I(a_n, g_n)$ es constante, y vale siempre $I(a, b)$. A su vez, para ello basta probar que se cumple $I(a, b) = I(a_1, g_1)$, ya que aplicando esto repetidamente obtenemos la igualdad para todo n . Nos ocupamos de ello a continuación. ■

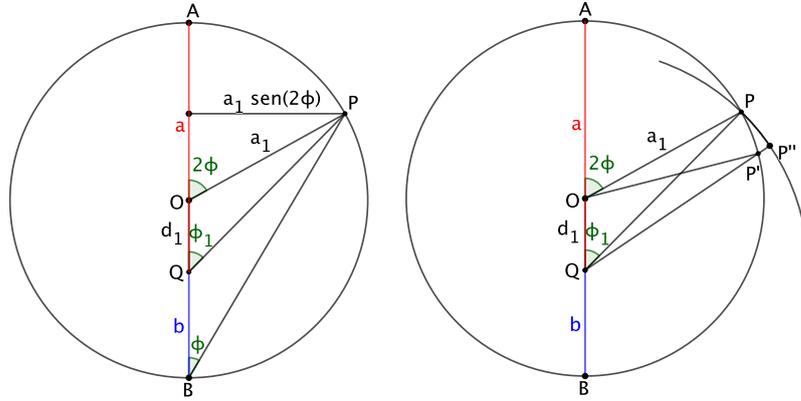
La transformación de Landen Gauss probó la igualdad que necesitamos para completar la prueba del teorema 6.17 aplicando a la integral el cambio de variable determinado por

$$\operatorname{sen} \phi = \frac{2a \operatorname{sen} \phi'}{a + b + (a - b) \operatorname{sen}^2 \phi'}.$$

Aquí vamos a emplear otro debido a Landen que nos permitirá extraer más consecuencias. El argumento no es el original de Landen, sino que es de Cayley.

Fijados dos números reales $a \geq b > 0$, consideramos una circunferencia C de radio $a_1 = (a + b)/2$, como muestra la figura de la izquierda. Sea Q el punto determinado por $\overline{AQ} = a$, $\overline{QB} = b$. Llamamos $d_1 = \overline{OQ} = (a - b)/2$.

Para cada ángulo $0 < \phi < \pi/2$, podemos considerar el punto P que muestra la figura de la izquierda y, a su vez, el ángulo ϕ_1 , que cumplirá $0 < \phi_1 < \pi$. Esto determina un cambio de variable $[0, \pi/2] \rightarrow [0, \pi]$.



Para obtener expresiones analíticas entre ϕ y ϕ_1 aplicamos el teorema del coseno al triángulo OPQ , de modo que

$$\overline{QP}^2 = \overline{OQ}^2 + \overline{OP}^2 + 2\overline{OQ}\overline{OP} \cos 2\phi,$$

o equivalentemente:

$$\overline{QP}^2 = d_1^2 + a_1^2 + 2d_1a_1 \cos 2\phi.$$

Ahora observamos que

$$d_1^2 + a_1^2 = \frac{a^2 + b^2 - 2ab}{4} + \frac{a^2 + b^2 + 2ab}{4} = \frac{a^2 + b^2}{2},$$

$$d_1a_1 = \frac{a-b}{2} \frac{a+b}{2} = \frac{a^2 - b^2}{4},$$

luego

$$\begin{aligned} \overline{QP}^2 &= \frac{a^2 + b^2}{2} + \frac{a^2 - b^2}{2} \cos 2\phi = \\ &= \frac{a^2 + b^2}{2} (\cos^2 \phi + \sin^2 \phi) + \frac{a^2 - b^2}{2} (\cos^2 \phi - \sin^2 \phi) = a^2 \cos^2 \phi + b^2 \sin^2 \phi. \end{aligned}$$

Por lo tanto:

$$\begin{aligned} \sin \phi_1 &= \frac{a_1 \sin 2\phi}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}}, \\ \cos \phi_1 &= \frac{d_1 + a_1 \cos 2\phi}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}}. \end{aligned} \quad (6.32)$$

Ahora consideremos un punto P' cercano a P sobre la circunferencia C , y sea P'' el punto en el que la semirrecta QP' corta a la circunferencia C^* de centro Q que pasa por P (véase la figura de la derecha). Vamos a analizar el triángulo $T = PP'P''$. Por la interpretación geométrica de la derivada, sabemos que cuando P' tiende a P , la recta PP' tiende a la tangente a C , es decir, a la recta perpendicular al radio OP . Más precisamente, la pendiente de PP' tiende a la

pendiente de dicha perpendicular. Similarmente, la pendiente de PP'' tiende a la pendiente de la tangente a C^* por P . Por consiguiente, el ángulo \hat{P} de T tiende al ángulo que forman ambas tangentes, que es el mismo que forman OP y QP , que es $\pi - \phi_1 - (\pi - 2\phi) = 2\phi - \phi_1$.

Similarmente, el ángulo \hat{P}'' tiende al ángulo que forma la tangente a C^* por P con OP , es decir, tiende a $\pi/2$. La trigonometría nos da entonces que

$$\lim_{\Delta\phi \rightarrow 0} \frac{\overline{PP''}}{\overline{PP'}} = \cos(2\phi - \phi_1),$$

donde $\Delta\phi = \phi' - \phi$ y ϕ' es el ángulo asociado a P' .

En efecto, por el teorema de los senos,

$$\frac{\overline{P'P''}}{\overline{PP'}} = \frac{\text{sen } \hat{P}}{\text{sen } \hat{P}''},$$

y por el teorema del coseno

$$\overline{P'P''}^2 = \overline{PP'}^2 + \overline{PP''}^2 - 2\overline{PP'}\overline{PP''}\cos\hat{P},$$

luego, llamando $X = \overline{PP''}/\overline{PP'}$, tenemos que

$$\frac{\text{sen}^2 \hat{P}}{\text{sen}^2 \hat{P}''} = 1 + X^2 - 2X \cos \hat{P} = 1 + (X - \cos \hat{P})^2 + \cos^2 \hat{P},$$

de donde se sigue que

$$\lim_{\Delta\phi \rightarrow 0} (X - \cos \hat{P})^2 = 1 - \text{sen}^2(2\phi - \phi_1) - \cos^2(2\phi - \phi_1) = 0,$$

luego $\lim_{\Delta\phi \rightarrow 0} X = \lim_{\Delta\phi \rightarrow 0} \cos \hat{P} = \cos(2\phi - \phi_1)$.

Por otra parte, las longitudes de los arcos cumplen

$$\widehat{PP'} = a_1 2\Delta\phi, \quad \widehat{PP''} = \overline{QP} \Delta\phi_1,$$

y las cuerdas correspondientes son

$$\overline{PP'} = 2a_1 \text{sen } \Delta\phi, \quad \overline{PP''} = 2\overline{QP} \text{sen } \frac{\Delta\phi_1}{2},$$

luego

$$\lim_{\Delta\phi \rightarrow 0} \frac{\overline{PP'}}{\widehat{PP'}} = \lim_{\Delta\phi \rightarrow 0} \frac{\text{sen } \Delta\phi}{\Delta\phi} = 1, \quad \lim_{\Delta\phi \rightarrow 0} \frac{\overline{PP''}}{\widehat{PP''}} = \lim_{\Delta\phi \rightarrow 0} \frac{\text{sen}(\Delta\phi_1/2)}{\Delta\phi_1/2} = 1.$$

Esto nos da que

$$\lim_{\Delta\phi \rightarrow 0} \frac{\overline{PP''}}{\overline{PP'}} = \lim_{\Delta\phi \rightarrow 0} \frac{\widehat{PP''} \widehat{PP'} \overline{PP''}}{\widehat{PP'} \overline{PP'} \widehat{PP''}} = \lim_{\Delta\phi \rightarrow 0} \frac{\overline{QP} \Delta\phi_1}{2a_1 \Delta\phi},$$

luego

$$\cos(2\phi - \phi_1) = \frac{\sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi}}{2a_1} \frac{d\phi_1}{d\phi}.$$

Finalmente, usando (6.32):

$$\begin{aligned} \cos(2\phi - \phi_1) &= \cos 2\phi \cos \phi_1 + \operatorname{sen} 2\phi \operatorname{sen} \phi_1 = \\ &= \frac{\cos 2\phi(d_1 + a_1 \cos 2\phi) + a_1 \operatorname{sen}^2 2\phi}{\sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi}} = \frac{d_1 \cos 2\phi + a_1}{\sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi}} \\ &= \frac{(a-b)(\cos^2 \phi - \operatorname{sen}^2 \phi) + a+b}{2\sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi}} = \frac{a \cos^2 \phi + b \operatorname{sen}^2 \phi}{\sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi}}, \end{aligned}$$

pero, por otra parte,

$$\begin{aligned} a_1^2 \cos^2 \phi_1 + g_1^2 \operatorname{sen}^2 \phi_1 &= a_1^2(1 - \operatorname{sen}^2 \phi_1) + g_1^2 \operatorname{sen}^2 \phi_1 \\ &= a_1^2 - \frac{(a_1^2 - g_1^2)a_1^2 \operatorname{sen}^2 2\phi}{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi} = \frac{a_1^2(a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi - (a_1^2 - g_1^2) \operatorname{sen}^2 2\phi)}{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi} \\ &= \frac{a_1^2(a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi - (a^2 + b^2 - 2ab) \cos^2 \phi \operatorname{sen}^2 \phi)}{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi} \\ &= \frac{a_1^2(a^2 \cos^4 \phi + b^2 \operatorname{sen}^4 \phi + 2ab \cos^2 \phi \operatorname{sen}^2 \phi)}{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi} \\ &= \frac{a_1^2(a \cos^2 \phi + b \operatorname{sen}^2 \phi)^2}{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi}, \end{aligned}$$

luego

$$\cos(2\phi - \phi_1) = \frac{1}{a_1} \sqrt{a_1 \cos^2 \phi + g_1 \operatorname{sen}^2 \phi}.$$

En definitiva obtenemos la relación

$$\sqrt{a_1 \cos^2 \phi + g_1 \operatorname{sen}^2 \phi} = \frac{1}{2} \sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi} \frac{d\phi_1}{d\phi} \quad (6.33)$$

que determina la derivada de la transformación de Landen. Al aplicarla a la integral $I(a, b)$ obtenemos que

$$\begin{aligned} I(a, b) &= \int_0^{\pi/2} \frac{d\phi}{\sqrt{a^2 \cos^2 \phi + b^2 \operatorname{sen}^2 \phi}} = \frac{1}{2} \int_0^\pi \frac{d\phi'}{\sqrt{a_1^2 \cos^2 \phi' + g_1^2 \operatorname{sen}^2 \phi'}} \\ &= \frac{1}{2} I(a_1, g_1) + \frac{1}{2} \int_{\pi/2}^\pi \frac{d\phi'}{\sqrt{a_1^2 \cos^2 \phi' + g_1^2 \operatorname{sen}^2 \phi'}} = I(a_1, g_1), \end{aligned}$$

pues el cambio $\phi' = \pi - \phi$ transforma la segunda integral en $I(a_1, g_1)$, y con esto completamos la demostración del teorema 6.17.

Veamos ahora el efecto de aplicar la transformación de Landen a la integral

$$J(a, b) = \int_0^{\pi/2} \sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi} d\phi = aE(m), \quad (6.34)$$

donde $m = 1 - (b/a)^2$ (esto está demostrado al principio de la sección 6.1).

Partimos de (6.32):

$$\sin^2 \phi_1 = \frac{4a_1^2 \sin^2 \phi \cos^2 \phi}{a^2 \cos^2 \phi + b^2 \sin^2 \phi}.$$

Por abreviar, vamos a llamar $X = a^2 \cos^2 \phi + b^2 \sin^2 \phi$. Entonces

$$X - a^2 = (b^2 - a^2) \sin^2 \phi, \quad X - b^2 = (a^2 - b^2) \cos^2 \phi,$$

luego

$$\begin{aligned} (X - a^2)(X - b^2) &= -(a^2 - b^2)^2 \sin^2 \phi \cos^2 \phi = \\ &= -(a + b)^2 (a - b)^2 \sin^2 \phi \cos^2 \phi = -4a_1^2 (a - b)^2 \sin^2 \phi \cos^2 \phi, \end{aligned}$$

luego

$$(X - a^2)(X - b^2) = -X(a - b)^2 \sin^2 \phi_1.$$

Equivalentemente:

$$X^2 + X(-a^2 - b^2) + (a - b)^2 \sin^2 \phi_1 + a^2 b^2 = 0.$$

Ahora observamos que

$$\begin{aligned} &-(a^2 + b^2) + (a - b)^2 \sin^2 \phi_1 = \\ &-(a^2 + b^2)(\cos^2 \phi_1 + \sin^2 \phi_1) + (a - b)^2 \sin^2 \phi_1 = \\ &-((a^2 + b^2) \cos^2 \phi_1 + 2ab \sin^2 \phi_1), \end{aligned}$$

luego

$$X^2 - X((a^2 + b^2) \cos^2 \phi_1 + 2ab \sin^2 \phi_1) + a^2 b^2 = 0.$$

El miembro izquierdo es:

$$\left(X - \frac{1}{2}(a^2 + b^2) \cos^2 \phi_1 + ab \sin^2 \phi_1\right)^2 - \left(\frac{1}{2}(a^2 + b^2) \cos^2 \phi_1 + ab \sin^2 \phi_1\right)^2 + a^2 b^2,$$

luego

$$\begin{aligned} &\left(X - \frac{1}{2}(a^2 + b^2) \cos^2 \phi_1 + ab \sin^2 \phi_1\right)^2 = \\ &\left(\frac{1}{2}(a^2 + b^2) \cos^2 \phi_1 + ab \sin^2 \phi_1\right)^2 - a^2 b^2 (\cos^2 \phi_1 + \sin^2 \phi_1)^2 = \\ &\frac{1}{4}(a^2 + b^2)^2 \cos^4 \phi_1 + ab(a^2 + b^2) \cos^2 \phi_1 \sin^2 \phi_1 + a^2 b^2 \sin^4 \phi_1 \\ &- a^2 b^2 \cos^4 \phi_1 - a^2 b^2 \sin^4 \phi_1 - 2a^2 b^2 \cos^2 \phi_1 \sin^2 \phi_1 = \end{aligned}$$

$$\begin{aligned} & \frac{1}{4}(a^2 - b^2)^2 \cos^4 \phi_1 + ab(a - b)^2 \cos^2 \phi_1 \sin^2 \phi_1 = \\ & (a - b)^2 \cos^2 \phi_1 \left(\frac{1}{4}(a + b)^2 \cos^2 \phi_1 + ab \sin^2 \phi_1 \right) = \\ & = 4d_1^2 \cos^2 \phi_1 (a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1). \end{aligned}$$

Así pues

$$\begin{aligned} X &= a^2 \cos^2 \phi + b^2 \sin^2 \phi = \frac{1}{2}(a^2 + b^2) \cos^2 \phi_1 \\ &+ ab \sin^2 \phi_1 + 2d_1 \cos \phi_1 \sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1} = \\ & \frac{1}{2}(a^2 + b^2) \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1 + 2d_1 \cos \phi_1 \sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1} = \\ & 2(a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1) - g_1^2 + 2d_1 \cos \phi_1 \sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1}. \end{aligned}$$

Equivalentemente:

$$\begin{aligned} & \frac{a^2 \cos^2 \phi + b^2 \sin^2 \phi}{\sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1}} = \\ & 2 \left(\sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1} - \frac{g_1^2/2}{\sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1}} + d_1 \cos \phi_1 \right). \end{aligned}$$

El primer miembro es

$$\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi} \frac{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}}{\sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1}},$$

por lo que (6.33) nos da la relación final:

$$\begin{aligned} & \sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi} d\phi = \\ & \left(\sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1} - \frac{g_1^2/2}{\sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1}} + d_1 \cos \phi_1 \right) d\phi_1. \end{aligned}$$

Por consiguiente, el cambio de variable $\phi(\phi_1)$ nos da que

$$J(a, b) = \int_0^\pi \sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1} d\phi_1 - \frac{g_1^2}{2} \int_0^\pi \frac{d\phi_1}{\sqrt{a_1^2 \cos^2 \phi_1 + g_1^2 \sin^2 \phi_1}}.$$

Descomponiendo cada integral en suma de las integrales en los intervalos $[0, \pi/2]$ y $[\pi/2, \pi]$ y aplicando el cambio $\phi_1 = \pi - \phi$ obtenemos que

$$J(a, b) = 2J(a_1, g_1) - g_1^2 I(a_1, g_1).$$

De aquí deducimos la fórmula siguiente:

Teorema 6.18 *Dados números reales $0 < b < a$, definimos las sucesiones*

$$a_0 = a, \quad g_0 = b, \quad d_0 = \sqrt{a^2 - b^2},$$

$$a_{n+1} = \frac{a_n + g_n}{2}, \quad g_{n+1} = \sqrt{a_n g_n}, \quad d_{n+1} = \frac{a_n - g_n}{2} = \sqrt{a_{n+1}^2 - g_{n+1}^2}.$$

Entonces

$$\int_0^{\pi/2} \sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi} d\phi =$$

$$\int_0^{\pi/2} \frac{d\phi}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}} \left(a^2 - \sum_{j=0}^{\infty} 2^{j-1} d_j^2 \right).$$

DEMOSTRACIÓN: Teniendo en cuenta que $I(a, b) = I(a_1, g_1)$, la relación que acabamos de obtener implica que

$$J(a_0, g_0) - a_0^2 I(a, b) = 2(J(a_1, g_1) - a_1^2 I(a, b)) + I(a, b)(2a_1^2 - a^2 - g_1^2).$$

Pero

$$2a_1^2 - a^2 - g_1^2 = -\frac{1}{2}(a^2 - b^2) = -\frac{1}{2}d_0^2,$$

luego

$$J(a_0, g_0) - a_0^2 I(a, b) = 2(J(a_1, g_1) - a_1^2 I(a, b)) - \frac{1}{2}d_0^2 I(a, b).$$

Aplicándolo a a_n y g_n obtenemos que

$$J(a_n, g_n) - a_n^2 I(a, b) = 2(J(a_{n+1}, g_{n+1}) - a_{n+1}^2 I(a, b)) - \frac{1}{2}d_n^2 I(a, b).$$

Razonando por inducción concluimos que

$$J(a, b) - a^2 I(a, b) = 2^n (J(a_n, g_n) - a_n^2 I(a, b)) - \frac{1}{2} I(a, b) \sum_{j=0}^{n-1} 2^j d_j^2.$$

Además,

$$J(a_n, g_n) - a_n^2 I(a, b) = \int_0^{\pi/2} \sqrt{a_n^2 \cos^2 \phi + g_n^2 \sin^2 \phi} d\phi$$

$$- \int_0^{\pi/2} \frac{a_n^2}{\sqrt{a_n^2 \cos^2 \phi + g_n^2 \sin^2 \phi}} d\phi = \int_0^{\pi/2} \frac{(g_n^2 - a_n^2) \sin^2 \phi}{\sqrt{a_n^2 \cos^2 \phi + g_n^2 \sin^2 \phi}} d\phi,$$

luego

$$2^n |J(a_n, g_n) - a_n^2 I(a, b)| \leq 2^n \int_0^{\pi/2} \frac{a_n^2 - g_n^2}{\sqrt{a_n^2 \cos^2 \phi + g_n^2 \sin^2 \phi}} d\phi$$

$$= 2^n (a_n^2 - g_n^2) I(a_n, g_n) = 2^n (a_n - g_n)(a_n + g_n) I(a, b) \leq K \frac{1}{2^{2^n - N - n}}$$

para todo n por encima de un cierto N , por (6.30).

Esto implica que

$$\lim_n 2^n (J(a_n, g - n) - a_n^2 I(a, b)) = 0,$$

luego

$$J(a, b) - a^2 I(a, b) = -\frac{1}{2} I(a, b) \sum_{j=0}^{\infty} 2^j d_j^n,$$

de donde se sigue inmediatamente la fórmula del enunciado. \blacksquare

Aunque en la demostración ha resultado más natural, en la práctica no es eficiente calcular d_{n+1} dividiendo $a_n - g - n$ entre 2 para luego multiplicar d_j^2 por 2^{j-1} . Si omitimos la división, hay que dividir $2^{j-1} d_j^2$ entre 4, luego tenemos el enunciado alternativo:

Teorema 6.19 *Dados números reales $0 < b < a$, definimos las sucesiones*

$$\begin{aligned} a_0 &= a, & g_0 &= b, & d_0 &= 2\sqrt{a^2 - b^2}, \\ a_{n+1} &= \frac{a_n + g_n}{2}, & g_{n+1} &= \sqrt{a_n g_n}, & d_{n+1} &= a_n - g_n. \end{aligned}$$

Entonces

$$\begin{aligned} &\int_0^{\pi/2} \sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi} d\phi = \\ &\int_0^{\pi/2} \frac{d\phi}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}} \left(a^2 - \sum_{j=0}^{\infty} 2^{j-3} d_j^2 \right). \end{aligned}$$

En términos de K y E , teniendo en cuenta que, según (6.31), (6.34),

$$J(a, b) = aE(m) \quad I(a, b) = K(m)/a, \quad m = 1 - (b/a)^2$$

y tomando $a_0 = 1$, $g_0 = \sqrt{1 - m}$, $d_0 = 2\sqrt{m}$, la relación es

$$E(m) = K(m) \left(1 - \sum_{j=0}^{\infty} 2^{j-3} d_j^2 \right).$$

Así pues, si $0 < m < 1$, una de las formas más rápidas de calcular las integrales elípticas de primera especie es mediante las sucesiones:

$$K(m) = I(1, \sqrt{1 - m}) = \frac{\pi}{2M(1, \sqrt{1 - m})} = \lim_n \frac{\pi}{2a_n} = \lim_n \frac{\pi}{2g_n},$$

mientras que las integrales elípticas de segunda especie pueden calcularse como

$$E(m) = J(1, \sqrt{1 - m}) = \frac{\pi}{2M(1, \sqrt{1 - m})} \left(1 - \sum_{j=0}^{\infty} 2^{j-3} d_j^2 \right) =$$

$$\lim_n \frac{\pi}{2a_n} \left(1 - \sum_{j=0}^n 2^{j-3} d_j^2 \right) = \lim_n \frac{\pi}{2b_n} \left(1 - \sum_{j=0}^n 2^{j-3} d_j^2 \right).$$

Ejemplo La tabla siguiente muestra el cálculo de $K(3/4)$ y $E(3/4)$:

n	g_n	a_n
0	0.50000000000000000000	1.00000000000000000000
1	0.70710678118654752440	0.75000000000000000000
2	0.72823765756098513043	0.72855339059327376220
3	0.72839550696977747310	0.72839552407712944631
4	0.72839551552345340948	0.72839551552345345971
5	0.72839551552345343459	0.72839551552345343459
n	$K(m)$	
0	1.5707963267948966192	3.1415926535897932385
1	2.0943951023931954923	2.2214414690791831235
2	2.1560483378105888854	2.1569831091347438635
3	2.1565156221753028830	2.1565156728239840340
4	2.1565156474996431611	2.1565156474996433098
5	2.1565156474996432354	2.1565156474996432354
n	$E(m)$	
0	0.98174770424681038702	1.9634954084936207740
1	1.1780972450961724644	1.2495608263570405070
2	1.2107938107304695835	1.2113187596909759615
3	1.2110560133468180814	1.2110560417901037433
4	1.2110560275684594831	1.2110560275684595666
5	1.2110560275684595248	1.2110560275684595248

En el ejemplo tras el teorema 6.2 calculamos $E(3/4)$ usando el desarrollo en serie de potencias. Vemos que la convergencia de las sucesiones que acabamos de obtener es mucho más rápida. Sólo con cinco iteraciones conseguimos 15 decimales exactos. ■

Uniendo varios de los resultados que hemos obtenido resulta una sucesión que converge a π muy rápidamente:

$$\pi = \frac{4M^2(1, 1/\sqrt{2})}{1 - \sum_{j=1}^{\infty} 2^{j-1} d_j^2}. \quad (6.35)$$

En efecto, tenemos que

$$K(1/2) = \frac{\pi}{2M(1, 1/\sqrt{2})}, \quad E(1/2) = \frac{\pi}{2M(1, 1/\sqrt{2})} \left(1 - \sum_{j=0}^{\infty} 2^{j-3} d_j^2\right),$$

donde la sucesión d_j viene dada por el teorema 6.19, de modo que en particular $d_0^2 = 2$. Por otra parte, la relación de Legendre 6.6 nos da que

$$2K(1/2)E(1/2) - K^2(1/2) = \frac{\pi}{2}.$$

Sustituyendo las expresiones precedentes y simplificando π , queda

$$\frac{2\pi}{4M^2(1, 1/\sqrt{2})} \left(1 - \sum_{j=0}^{\infty} 2^{j-3} d_j^2\right) - \frac{\pi}{4M^2(1, 1/\sqrt{2})} = \frac{1}{2},$$

de donde

$$\pi \left(2 - \sum_{j=0}^{\infty} 2^{j-1} d_j^2 \right) = 4M^2(1, 1/\sqrt{2})$$

y, como $d_0^2 = 2$, el primer término de la suma vale 1 y llegamos a (6.35). Así pues:

Teorema 6.20 Si $a_0 = 1$, $g_0 = 1/\sqrt{2}$,

$$a_{n+1} = \frac{a_n + g_n}{2}, \quad g_{n+1} = \sqrt{a_n g_n}, \quad d_{n+1} = a_n - g_n,$$

entonces

$$\lim_n \frac{4a_n^2}{1 - \sum_{j=1}^n 2^{j-1} d_j^2} = \pi.$$

La tabla siguiente ilustra muestra la velocidad de la convergencia: a la cuarta iteración obtenemos ya 20 cifras decimales exactas y a la quinta 42:

a_1	0.85355339059327376220042218105242451964241796884424
g_1	0.84089641525371454303112547623321489504003426235678
s_1	0.08578643762690495119831127579030192143032812462305
x_1	3.1876726427121086272019299705253692326510535718594
a_2	0.84722490292349415261577382864281970734122611560051
g_2	0.84720126674689146040363145369335239796398101361200
s_2	0.08610683567639737162109343907794333391179532051109
x_2	3.1416802932976532939180704245600093827957194388154
a_3	0.84721308483519280650970264116808605265260356460626
g_3	0.84721308475276536670429805177990207039211065605945
s_3	0.08610683791107274919571249224164455099320272044588
x_3	3.1415926538954464960029147588180434861088792372613
a_4	0.84721308479397908660700034647399406152235711033285
g_4	0.84721308479397908660599790049038921144053485858626
s_4	0.08610683791107274925006675490463332677644846160314
x_4	3.1415926535897932384663606027066313217577024113424
a_5	0.84721308479397908660649912348219163648144598445956
g_5	0.84721308479397908660649912348219163648144583619433
s_5	0.08610683791107274925006675490463332677646453997034
x_5	3.1415926535897932384626433832795028841971699491647

Capítulo VII

Introducción al análisis complejo

En la definición de derivada 1.1 consideramos el caso de funciones definidas en un abierto del plano complejo \mathbb{C} , y en la sección 1.2 vimos que las reglas básicas para el cálculo de derivadas valen igualmente para funciones de variable compleja. Sin embargo, aparte de haber probado estos hechos básicos, hemos trabajado exclusivamente con funciones de variable real. En este último capítulo presentamos los aspectos básicos del cálculo diferencial e integral con funciones de variable compleja. Veremos que éste dista mucho de ser una mera generalización del caso real, sino que involucra ideas nuevas que le dan un carácter muy diferente. Más concretamente, veremos que este “carácter” deriva en última instancia de un teorema nada trivial debido a Cauchy con el que, entre otras cosas, justificaremos el cálculo que hizo Euler de las integrales de Fresnel discutidas en la introducción a raíz de su estudio de la espiral de Euler. Primeramente derivaremos sus ecuaciones:

La clotoide Del mismo modo que la elástica rectangular está determinada por el hecho de que su curvatura es proporcional a su altura, la *clotoide*, o *espiral de Euler*, está determinada por que su curvatura es proporcional a su longitud (compárese con el hecho de que —según hemos visto— la curvatura de la espiral de Bernoulli es inversamente proporcional a su longitud).

Si $\vec{r}(s) = (x(s), y(s))$ es una curva parametrizada por el arco, entonces el vector $\vec{v} = (x'(s), y'(s))$ tiene norma 1, luego coincide con el vector tangente unitario $\vec{r}'(s) = (\cos \phi(s), \sin \phi(s))$, luego

$$\vec{r}'' = (-\sin \phi, \cos \phi)\phi',$$

luego, por la propia definición de curvatura,

$$\kappa = \frac{\|\vec{r}''\|}{v} = \phi'.$$

Que la curvatura sea proporcional a la longitud significa (midiendo la longitud desde el punto de curvatura 0) que se cumple la relación $s = a^2\kappa$, donde hemos llamado a^2 al factor de proporcionalidad. Equivalentemente, $s = a^2\phi'$, de donde

$$\phi = \frac{s^2}{2a^2},$$

donde hemos introducido la condición $\phi(0) = 0$, que significa que, en el punto de curvatura 0, la tangente es horizontal. Esto no es ninguna restricción, sino que puede conseguirse sin más que girar la curva. Así pues, tenemos que

$$(x', y') = \left(\cos \frac{s^2}{2a^2}, \sin \frac{s^2}{2a^2} \right),$$

luego integrando llegamos a que

$$x = \int_0^s \cos \frac{u^2}{2a^2} du, \quad y = \int_0^s \sin \frac{u^2}{2a^2} du.$$

El cambio de variable $u = \sqrt{2a}t$ nos da

$$x = \sqrt{2a} \int_0^{s/\sqrt{2a}} \cos t^2 dt, \quad y = \sqrt{2a} \int_0^{s/\sqrt{2a}} \sin t^2 dt.$$

Ésta es la parametrización por el arco de la clotoide, cuya forma es la que se muestra en la introducción, y el polo alrededor del cual “se enrolla” es

$$\sqrt{2a} \left(\int_0^{+\infty} \cos t^2 dt, \int_0^{+\infty} \sin t^2 dt \right).$$

En este capítulo probaremos que las integrales valen $\sqrt{\pi/8}$, con lo que el polo de la espiral es

$$\left(\frac{\sqrt{\pi}}{2}a, \frac{\sqrt{\pi}}{2}a \right).$$

Pero para llegar a esta conclusión necesitamos adentrarnos en el cálculo diferencial e integral con funciones de variable compleja.

7.1 Funciones holomorfas

Consideremos una función $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ definida sobre un abierto A en el plano complejo. En 1.1 hemos definido lo que significa que f sea derivable en un punto $z \in A$. Si f es derivable en todos los puntos de A se dice que f es una función *holomorfa*¹ en A . Las funciones holomorfas en \mathbb{C} se llaman funciones *enteras*.

¹El término fue introducido en 1875 por Charles Briot y Jean-Claude Bouquet, dos alumnos de Cauchy, y significa “de forma entera”, porque antiguamente se llamaba “funciones enteras” a los polinomios, y las funciones holomorfas presentan muchas propiedades análogas las de los polinomios.

Notemos que “holomorfa” no es sinónimo de “derivable” porque la derivabilidad tiene sentido en un punto, mientras que la holomorfía exige la derivabilidad en todos los puntos del abierto. Cuando se habla de una función holomorfa en un punto z_0 , hay que entenderlo como que la función es holomorfa en un abierto que contiene a z_0 .

El teorema 1.5 (que hemos probado para funciones reales y complejas) muestra que la suma y el producto de funciones holomorfas en un abierto A es también una función holomorfa, así como el cociente si el denominador no se anula. En particular, todo polinomio es una función entera y, por ejemplo, la función $1/z$ es holomorfa en $\mathbb{C} \setminus \{0\}$, pues en este abierto es cociente de funciones holomorfas que no se anulan.

En la sección 1.2 hemos visto también que las funciones e^z , $\operatorname{sen} z$ y $\operatorname{cos} z$ son enteras, y que cualquier logaritmo continuo $\log z$ definido en un abierto de \mathbb{C} es una función holomorfa con derivada $1/z$.

Tenemos probado que las reglas de derivación de sumas, productos, cocientes, así como la regla de la cadena 1.6 se pueden aplicar igualmente a funciones de variable compleja. Por ejemplo, la derivada de $\operatorname{sen} z^3$ es $3z^2 \operatorname{cos} z^3$, etc.

Cada número complejo $z = x + yi \in \mathbb{C}$ puede identificarse con el par (x, y) de números reales, por lo que podemos considerar a la función f como una función $f : A \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$. Esto nos permitirá relacionar el concepto de derivación compleja con la teoría sobre funciones diferenciables que hemos estudiado en el capítulo V. Por ejemplo, la función $f(z) = z^2$, definida en $A = \mathbb{C}$, puede expresarse como

$$f(x, y) = (x + yi)^2 = x^2 - y^2 + 2xyi = (x^2 - y^2, 2xy).$$

Las funciones coordenadas que en la definición 5.14 llamábamos f_1 y f_2 son en este contexto la *parte real* y *parte imaginaria* de la función f , que representamos por $\operatorname{Re} f, \operatorname{Im} f : A \subset \mathbb{R}^2 \rightarrow \mathbb{R}$. En nuestro ejemplo son $\operatorname{Re} f(x, y) = x^2 - y^2$, $\operatorname{Im} f(x, y) = 2xy$. Es claro entonces que

$$df(x, y) = (2x \, dx - 2y \, dy, 2y \, dx + 2x \, dy)$$

o, más explícitamente,

$$df(x, y)(v) = (2xv_1 - 2yv_2, 2yv_1 + 2xv_2).$$

Si expresamos esto en términos de números complejos, tenemos que

$$df(z)(v) = 2xv_1 - 2yv_2 + (2yv_1 + 2xv_2)i = (2x + 2yi)(v_1 + v_2i) = 2zv = f'(z)v.$$

Vemos así que $df(z)$, que en principio es una aplicación lineal $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, no es cualquier aplicación lineal, sino la que se calcula sobre un $v \in \mathbb{R}^2$ multiplicándolo por $f'(z) = 2z$. No todas las aplicaciones lineales de \mathbb{R}^2 en sí mismo pueden expresarse como el producto por un número complejo fijo, y resulta que esto caracteriza a las funciones derivables en sentido complejo:

Teorema 7.1 Una aplicación $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ definida en un abierto A , es derivable en un punto $z \in A$ si y sólo si, vista como $f : A \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$, es diferenciable en z y su diferencial es de la forma $df(z)(v) = cv$, para cierto $c \in \mathbb{C}$, y en tal caso $f'(z) = c$.

DEMOSTRACIÓN: Si f es derivable en z , llamemos $c = f'(z) = a + bi$. Según la definición 1.1, esto significa que existe

$$\lim_{v \rightarrow 0} \frac{f(z+v) - f(z)}{v} = c$$

o, equivalentemente,

$$\lim_{v \rightarrow 0} \frac{f(z+v) - f(z) - cv}{v} = 0.$$

Aquí hay que tener en cuenta que $v = v_1 + v_2i = (v_1, v_2)$ es un número complejo, cuyo valor absoluto como número complejo coincide con su norma como elemento de \mathbb{R}^2 , es decir, $|v| = \|v\| = \sqrt{v_1^2 + v_2^2}$. El cociente $v/\|v\|$ tiene norma 1, luego está acotado, y el producto de una función acotada por otra que tiende a 0, tiende también a 0, luego, multiplicando la igualdad anterior por $v/\|v\|$ obtenemos que

$$\lim_{v \rightarrow 0} \frac{f(z+v) - f(z) - cv}{\|v\|} = 0.$$

Si llamamos $\phi : \mathbb{C} \rightarrow \mathbb{C}$ a la función $\phi(v) = cv$, tenemos que

$$\lim_{v \rightarrow 0} \frac{f(z+v) - f(z) - \phi(v)}{\|v\|} = 0. \quad (7.1)$$

Notemos que ϕ es una aplicación lineal. Concretamente,

$$\phi(v) = (a + bi)(v_1 + v_2i) = av_1 - bv_2 + (av_2 + bv_1)i = (av_1 - bv_2, bv_1 + av_2).$$

El límite anterior equivale a los límites

$$\lim_{v \rightarrow 0} \frac{f_1(z+v) - f_1(z) - \phi_1(v)}{\|v\|} = 0, \quad \lim_{v \rightarrow 0} \frac{f_2(z+v) - f_2(z) - \phi_2(v)}{\|v\|} = 0 \quad (7.2)$$

que afirman que las funciones $f_1 = \operatorname{Re} f$ y $f_2 = \operatorname{Im} f$ son diferenciables en z en el sentido de la definición 5.1, con diferenciales ϕ_1 y ϕ_2 , lo cual a su vez implica que f es diferenciable en z en el sentido de la definición 5.14, y que su diferencial es

$$df(z)(v) = \phi(v) = f'(z)v.$$

Recíprocamente, si f es diferenciable en z y $df(z)(v) = cv$, para cierto número complejo $c = a + bi \in \mathbb{C}$, llamando $\phi(v) = cv$, tenemos que f_1 y f_2 son diferenciables en z con diferenciales ϕ_1 y ϕ_2 , lo cual equivale a que existen los límites (7.2), lo cual a su vez equivale a que existen los límites (7.1), y prosiguiendo el razonamiento anterior en sentido inverso llegamos a que existe $f'(z) = c$. ■

Como ya hemos señalado, no todas las aplicaciones lineales $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ son de la forma $\phi(v) = cv$, para cierto $c \in \mathbb{C}$. Tal y como hemos visto en la demostración del teorema anterior, para que ϕ sea de esta forma, con $c = a + bi$, su expresión tiene que ser

$$\phi(v) = (av_1 - bv_2, bv_1 + av_2)$$

o, equivalentemente, su matriz asociada tiene que ser de la forma

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

(mientras que los coeficientes de una aplicación lineal arbitraria son cuatro números reales arbitrarios). Ahora, si una función $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ es diferenciable en un punto $z \in A$, la matriz de $df(z)$ es la *matriz jacobiana*

$$Jf = \begin{pmatrix} \frac{\partial \operatorname{Re} f}{\partial x} & \frac{\partial \operatorname{Im} f}{\partial x} \\ \frac{\partial \operatorname{Re} f}{\partial y} & \frac{\partial \operatorname{Im} f}{\partial y} \end{pmatrix},$$

luego el teorema anterior puede reformularse como sigue:

Teorema 7.2 Una función $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ definida en un abierto A es derivable en un punto $z \in A$ si y sólo si es diferenciable en z y satisface en dicho punto las ecuaciones de Cauchy-Riemann:

$$\frac{\partial \operatorname{Re} f}{\partial x} = \frac{\partial \operatorname{Im} f}{\partial y}, \quad \frac{\partial \operatorname{Re} f}{\partial y} = -\frac{\partial \operatorname{Im} f}{\partial x}.$$

En tal caso

$$f'(z) = \frac{\partial f}{\partial x}.$$

En la última igualdad hay que entender que

$$\begin{aligned} \frac{\partial f}{\partial x} &= \left(\frac{\partial \operatorname{Re} f}{\partial x}, \frac{\partial \operatorname{Im} f}{\partial x} \right) = \frac{\partial \operatorname{Re} f}{\partial x} + \frac{\partial \operatorname{Im} f}{\partial x} i, \\ \frac{\partial f}{\partial y} &= \left(\frac{\partial \operatorname{Re} f}{\partial y}, \frac{\partial \operatorname{Im} f}{\partial y} \right) = \frac{\partial \operatorname{Re} f}{\partial y} + \frac{\partial \operatorname{Im} f}{\partial y} i. \end{aligned}$$

Con esta notación, las ecuaciones de Cauchy-Riemann se pueden expresar en la forma

$$\frac{\partial f}{\partial y} = i \frac{\partial f}{\partial x}.$$

Ejemplo: La conjugación compleja La conjugación compleja es la aplicación $f : \mathbb{C} \rightarrow \mathbb{C}$ dada por $f(x, y) = (x, -y)$, luego

$$\frac{\partial \operatorname{Re} f}{\partial x} = 1, \quad \frac{\partial \operatorname{Re} f}{\partial y} = 0, \quad \frac{\partial \operatorname{Im} f}{\partial x} = 0, \quad \frac{\partial \operatorname{Im} f}{\partial y} = -1.$$

Claramente, no se cumplen las ecuaciones de Cauchy-Riemann, luego la conjugación no es derivable en ningún punto. ■

Ejercicio: Probar que las funciones $\operatorname{Re}(x, y) = x$, $\operatorname{Im}(x, y) = y$ no cumplen las ecuaciones de Cauchy-Riemann, luego no son derivables en ningún punto.

Ejemplo La función $f(z) = |z|^2 = x^2 + y^2$ sólo es derivable en 0. En efecto, es diferenciable en todo \mathbb{R}^2 , pero

$$\frac{\partial \operatorname{Re} f}{\partial x} = 2x, \quad \frac{\partial \operatorname{Re} f}{\partial y} = 2y, \quad \frac{\partial \operatorname{Im} f}{\partial x} = 0, \quad \frac{\partial \operatorname{Im} f}{\partial y} = 0,$$

luego las ecuaciones de Cauchy-Riemann sólo se cumplen en $(0, 0)$. Así pues, $f(z)$ es derivable en $z = 0$, pero no es holomorfa en dicho punto, ya que no es holomorfa en ningún abierto que lo contenga. ■

Interpretación geométrica de la derivada La interpretación geométrica de la derivación compleja no es tan relevante como en el caso real. De la propia definición se sigue que si una función f es derivable en un punto z_0 , entonces, para todo v suficientemente pequeño,

$$f(z_0 + v) \approx f(z_0) + f'(z_0)v$$

o, llamando $z = z_0 + v$, para todo z suficientemente próximo a z_0 se cumple que

$$f(z) \approx f(z_0) + f'(z_0)(z - z_0).$$

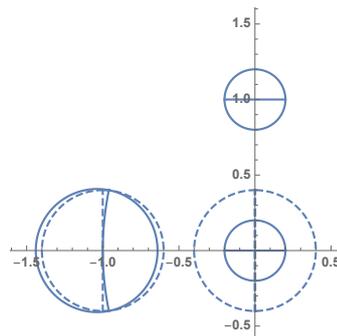
Consideremos por ejemplo $f(z) = z^2$ y $z_0 = i$, de modo que

$$f(z_0) = -1, \quad f'(z_0) = 2i.$$

La figura muestra una circunferencia centrada en z_0 con radio 0.2 y uno de sus diámetros. Al aplicarle $z \mapsto z - z_0$ obtenemos una circunferencia del mismo radio centrada en 0 y su diámetro horizontal. El efecto de multiplicar un número complejo por $f'(z_0) = 2i$ es duplicar su módulo y sumar $\pi/2$ a su argumento, es decir, girarlo 90° . Así pues, al aplicar $z \mapsto 2iz$ a la circunferencia de centro 0 obtenemos una circunferencia de radio doble y su diámetro vertical en lugar del horizontal, y si finalmente aplicamos $z \mapsto f(z_0) + z$ obtenemos una circunferencia del mismo radio con centro -1 (y su diámetro vertical).

La figura muestra estas transformaciones y también el resultado de aplicar $z \mapsto z^2$ a la circunferencia y el diámetro iniciales. Vemos que el resultado es aproximadamente el mismo, y la aproximación sería mucho mayor si tomáramos radios cada vez menores.

En general, el efecto de aplicar f a puntos z cercanos a z_0 es aproximadamente el que resulta de girar z un ángulo igual al argumento de $f'(z_0)$, aplicar una homotecia de centro z_0 y radio $|f'(z_0)|$ y trasladar el resultado sumándole $f(z_0)$. ■



7.2 La integral curvilínea

Generalizamos ahora el concepto de integral para funciones de variable compleja. Vamos a definir una integral análoga a la integral sobre una curva que hemos usado para definir el concepto físico de “trabajo”. Previamente conviene definir la integral de una función con valores complejos definida en un intervalo real:

Definición 7.3 Sea $h : [a, b] \rightarrow \mathbb{C}$ una función continua. Definimos

$$\int_a^b h(t) dt = \int_a^b \operatorname{Re} h(t) dt + i \int_a^b \operatorname{Im} h(t) dt.$$

Notemos que si $h : [a, b] \rightarrow \mathbb{R}$ la definición anterior se reduce a la integral usual de una función real. El teorema siguiente recoge las propiedades básicas de las integrales complejas.

Teorema 7.4 *Se cumple:*

a) Si $h_1, h_2 : [a, b] \rightarrow \mathbb{C}$ son continuas y $\alpha_1, \alpha_2 \in \mathbb{C}$, entonces

$$\int_a^b (\alpha_1 h_1(t) + \alpha_2 h_2(t)) dt = \alpha_1 \int_a^b h_1(t) dt + \alpha_2 \int_a^b h_2(t) dt.$$

b) Si $h : [a, b] \rightarrow \mathbb{C}$ es continua y $a < c < b$, entonces

$$\int_a^b h(t) dt = \int_a^c h(t) dt + \int_c^b h(t) dt.$$

c) Si $h : [c, d] \rightarrow \mathbb{C}$ es continua y $\rho : [a, b] \rightarrow [c, d]$ es biyectiva, derivable y con derivada continua, entonces

$$\int_{\rho(a)}^{\rho(b)} h(t) dt = \int_a^b h(\rho(s)) \rho'(s) ds,$$

con el convenio de que invertir los límites de integración equivale a cambiar el signo a la integral.

d) Si $h : [a, b] \rightarrow \mathbb{C}$ es continua, entonces

$$\left| \int_a^b h(t) dt \right| \leq \int_a^b |h(t)| dt.$$

DEMOSTRACIÓN: Para probar las tres primeras afirmaciones se separa la parte real de la parte imaginaria según la definición de integral compleja, se aplica la propiedad correspondiente de integrales reales y se vuelven a agrupar los sumandos. Veamos la cuarta propiedad.

Si $\int_a^b h(t) dt = 0$ el resultado es obvio. En otro caso sea

$$\alpha = \frac{\left| \int_a^b h(t) dt \right|}{\int_a^b h(t) dt} \in \mathbb{C}.$$

Así

$$\left| \int_a^b h(t) dt \right| = \alpha \int_a^b h(t) dt = \int_a^b \alpha h(t) dt = \int_a^b \operatorname{Re}(\alpha h(t)) dt + i \int_a^b \operatorname{Im}(\alpha h(t)) dt,$$

pero como se trata de un número real, la segunda parte ha de ser nula, luego

$$\left| \int_a^b h(t) dt \right| = \int_a^b \operatorname{Re}(\alpha h(t)) dt \leq \int_a^b |\alpha h(t)| dt = \int_a^b |h(t)| dt$$

donde hemos usado que, en general, $\operatorname{Re} z \leq |z|$, así como que $|\alpha| = 1$. ■

Ahora ya podemos introducir la integral curvilínea que necesitamos:

Definición 7.5 Sea $\phi : [a, b] \rightarrow \mathbb{C}$ un arco derivable y $f : \phi^* \rightarrow \mathbb{C}$ una función continua (donde $\phi^* \subset \mathbb{C}$ es la imagen de ϕ). Definimos la *integral curvilínea* de f a lo largo de ϕ como

$$\int_{\phi} f(\zeta) d\zeta = \int_a^b f(\phi(t)) \phi'(t) dt.$$

Antes de ver ejemplos conviene probar algunos hechos elementales:

Teorema 7.6 La integral curvilínea cumple las propiedades siguientes:

a) Sea ϕ un arco, sean $f, g : \phi^* \rightarrow \mathbb{C}$ continuas y $\alpha, \beta \in \mathbb{C}$. Entonces

$$\int_{\phi} (\alpha f(\zeta) + \beta g(\zeta)) d\zeta = \alpha \int_{\phi} f(\zeta) d\zeta + \beta \int_{\phi} g(\zeta) d\zeta.$$

b) Si el arco ψ se obtiene del arco ϕ mediante un cambio de parámetro² y $f : \phi^* \rightarrow \mathbb{C}$ es continua, entonces

$$\int_{\phi} f(\zeta) d\zeta = \pm \int_{\psi} f(\zeta) d\zeta,$$

donde el signo es el de la derivada del cambio de parámetro.

c) Si ϕ es un arco de longitud $L(\phi)$, $f : \phi^* \rightarrow \mathbb{C}$ es continua y $|f(z)| \leq M$, para todo $z \in \phi^*$, entonces

$$\left| \int_{\phi} f(\zeta) d\zeta \right| \leq L(\phi)M.$$

²Véase la definición 2.15.

DEMOSTRACIÓN: a) se sigue inmediatamente de la definición.

b) Sean $\phi : [a, b] \rightarrow \mathbb{C}$ y $\psi : [c, d] \rightarrow \mathbb{C}$. Sea $\rho : [a, b] \rightarrow [c, d]$ creciente biyectiva, derivable y con derivada continua tal que $\phi(t) = \psi(\rho(t))$. Entonces

$$\begin{aligned} \int_{\psi} f(\zeta) d\zeta &= \int_c^d f(\psi(s))\psi'(s) ds = \int_a^b f(\psi(\rho(t)))\psi'(\rho(t))\rho'(t) dt \\ &= \int_a^b f(\phi(t))\phi'(t) dt = \int_{\phi} f(\zeta) d\zeta. \end{aligned}$$

Si ρ es decreciente (es decir, si tiene derivada negativa), al hacer el cambio de variable obtenemos la integral de b hasta a , y al invertir los extremos aparece un signo negativo.

c) Sea $\phi : [a, b] \rightarrow \mathbb{C}$. Entonces

$$\begin{aligned} \left| \int_{\phi} f(\zeta) d\zeta \right| &= \left| \int_a^b f(\phi(s))\phi'(s) ds \right| \leq \int_a^b |f(\phi(s))| |\phi'(s)| ds \\ &\leq M \int_a^b |\phi'(s)| ds = L(\phi) M. \quad \blacksquare \end{aligned}$$

Conviene destacar la propiedad b): la integral curvilínea no depende realmente de la parametrización del arco sobre el que integramos (pero sí de su sentido, puesto que parametrizaciones con sentidos opuestos dan lugar a integrales con signos opuestos).

Ejemplo: Segmentos Dados dos números complejos z_0 y z_1 (no necesariamente distintos), definimos el *segmento* $[z_0, z_1] : [0, 1] \rightarrow \mathbb{C}$ mediante

$$[z_0, z_1](t) = (1-t)z_0 + tz_1.$$

Así:

$$\int_{[z_0, z_1]} f(\zeta) d\zeta = \int_0^1 f((1-t)z_0 + tz_1)(z_1 - z_0) dt.$$

En particular, si $z_0 < z_1$ son números reales, separando la parte real y al imaginaria y aplicando el cambio de variable $x = (1-t)z_0 + tz_1$ obtenemos que

$$\int_{[z_0, z_1]} f(\zeta) d\zeta = \int_{z_0}^{z_1} f(x) dx,$$

y si f toma únicamente valores reales, se trata de la integral usual que ya teníamos definida. Así pues, la integral curvilínea generaliza a la integral usual de funciones de variable real.

En general, el cambio de parámetro $t' = 1-t$ transforma el segmento $[z_0, z_1]$ en $[z_1, z_0]$, luego el apartado b) del teorema anterior implica que

$$\int_{[z_0, z_1]} f(\zeta) d\zeta = - \int_{[z_1, z_0]} f(\zeta) d\zeta.$$

Observemos también que la longitud de un segmento es, como cabía esperar:

$$L([z_0, z_1]) = \int_0^1 |[z_0, z_1]'| dt = \int_0^1 |z_1 - z_0| dt = |z_1 - z_0|.$$

Veamos un ejemplo concreto de cálculo de una integral sobre un segmento:

$$\begin{aligned} \int_{[1, i]} \zeta^2 d\zeta &= \int_0^1 (1-t+ti)^2(i-1) dt = \int_0^1 (-1+2t^2) dt + i \int_0^1 (1-4t+2t^2) dt \\ &= \left[-t + \frac{2}{3}t^3\right]_0^1 + i \left[t - 2t^2 + \frac{2}{3}t^3\right]_0^1 = -\frac{1}{3} - \frac{1}{3}i = -\frac{1+i}{3}. \end{aligned}$$

Hemos hecho este cálculo para ilustrar las definiciones, pero a continuación vamos a ver que la regla de Barrow es válida para integrales curvilíneas, de modo que la integral puede calcularse así:

$$\int_{[1, i]} \zeta^2 d\zeta = \left[\frac{\zeta^3}{3}\right]_1^i = \frac{-i}{3} - \frac{1}{3} = -\frac{1+i}{3}. \quad \blacksquare$$

Para probar la regla de Barrow conviene demostrar antes lo siguiente:

Teorema 7.7 Sea $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ una función holomorfa en un abierto A que tenga primitiva $F : A \subset \mathbb{C} \rightarrow \mathbb{C}$, es decir, que F es una función holomorfa tal que $F' = f$. Si $\phi : [a, b] \rightarrow A$ es un arco, entonces la función $\psi : [a, b] \rightarrow \mathbb{C}$ dada por $\psi(t) = F(\phi(t))$ también es un arco con derivada $\psi'(t) = f(\phi(t))\phi'(t)$.

DEMOSTRACIÓN: Vamos a aplicar la regla de la cadena 5.12 a ϕ y a las funciones coordenadas F_1 y F_2 de F . La conclusión es que las funciones coordenadas $\psi_1, \psi_2 :]a, b[\rightarrow \mathbb{R}$ son diferenciables y

$$d\psi_j(t)(v) = dF_j(\phi(t))(d\phi_1(t)(v), d\phi_2(t)(v)).$$

Notemos que, como ϕ_j es una función derivable, su diferencial es simplemente $d\phi_j(t)(v) = \phi_j'(t)v$, luego $\phi_j'(t) = d\phi_j(t)(1)$. Igualmente, que ψ_j sea diferenciable equivale a que sea derivable, y se cumple la misma relación, luego, para $v = 1$, la igualdad anterior se reduce a

$$\psi_j'(t) = dF_j(\phi(t))(\phi'(t)).$$

Combinando estas igualdades para $j = 1, 2$ queda

$$\psi'(t) = dF(\phi(t))(\phi'(t)).$$

Por último, como F es holomorfa, $dF(\phi(t))$ es la aplicación que actúa multiplicando por $F'(\phi(t)) = f(\phi(t))$, luego la igualdad anterior equivale a la del enunciado. \blacksquare

En particular, este teorema afirma que los arcos definidos por primitivas de funciones holomorfas se pueden derivar con las reglas usuales de derivación de funciones. Por ejemplo, si $\psi(t) = \sin t^2$, podemos aplicar el teorema anterior a la función $F(z) = \sin z^2$ y al arco $\phi(t) = t$, con lo que $\phi'(t) = 1$, y la conclusión es que $\psi'(t) = 2t \cos t^2$.

Teorema 7.8 (Regla de Barrow) Sea $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ una función holomorfa en un abierto A que tenga primitiva $F : A \subset \mathbb{C} \rightarrow \mathbb{C}$, es decir, que F es una función holomorfa tal que $F' = f$. Si $\phi : [a, b] \rightarrow A$ es un arco, entonces

$$\int_{\phi} f(\zeta) d\zeta = F(\phi(b)) - F(\phi(a)).$$

En particular, si ϕ es un arco cerrado (es decir, si $\phi(a) = \phi(b)$), entonces

$$\int_{\phi} f(\zeta) d\zeta = 0.$$

DEMOSTRACIÓN: Por definición

$$\int_{\phi} f(\zeta) d\zeta = \int_a^b f(\phi(t)) \phi'(t) dt,$$

y por el teorema anterior el integrando es la derivada de $F(\phi(t))$, que es continua en $[a, b]$, porque lo es ϕ y F también es continua en la imagen de ϕ , luego separando la parte real de la imaginaria y aplicando la regla de Barrow para integrales reales, obtenemos la conclusión. ■

Ejemplo: Circunferencias Es costumbre representar mediante

$$\int_{|z-z_0|=r} f(\zeta) d\zeta$$

la integral sobre el arco cerrado $\phi : [0, 2\pi] \rightarrow \mathbb{C}$ dado por $\phi(t) = z_0 + re^{it}$.

En realidad, como la parametrización es irrelevante, podemos tomar también el arco $\phi : [0, 1] \rightarrow \mathbb{C}$ dado por $\phi(t) = z_0 + re^{2\pi it}$, o cualquier otro que recorra en sentido positivo la circunferencia de centro z_0 y radio r . Explícitamente:

$$\int_{|z-z_0|=r} f(\zeta) d\zeta = ir \int_0^{2\pi} f(z_0 + e^{it}) e^{it} dt.$$

Por ejemplo

$$\int_{|z-z_0|=r} \frac{1}{\zeta - z_0} d\zeta = \int_0^{2\pi} \frac{1}{re^{it}} re^{it} i dt = i \int_0^{2\pi} dt = 2\pi i. \quad (7.3)$$

Este ejemplo es interesante porque muestra que la función continua $1/(z - z_0)$ no tiene primitiva en $\mathbb{C} \setminus \{z_0\}$, ya que si la tuviera, la integral tendría que anularse, por la regla de Barrow. ■

Es interesante entender por qué la función $1/(z - z_0)$ no tiene primitiva en $\mathbb{C} \setminus \{z_0\}$. Por simplificar vamos a tomar $z_0 = 0$.

Tras el teorema de la función inversa 1.8 hemos visto que existen ramas uniformes holomorfas del logaritmo $\log_{\alpha} : H_{\alpha} \rightarrow \mathbb{C}$, donde H_{α} es el plano complejo menos la semirrecta formada por el 0 y los puntos de argumento α , de modo que $\log_{\alpha} z$ es el único logaritmo de z con parte imaginaria en $]\alpha, \alpha + 2\pi[$.

Además, $\log'_\alpha(z) = 1/z$, de modo que la función $1/z$ tiene primitiva en cualquier abierto que resulte de quitarle al plano complejo una semirrecta arbitraria de origen en 0, pero si los puntos de dicha semirrecta tienen argumento α , entonces, cuando nos aproximamos a uno de sus puntos (no nulo), la parte imaginaria de la primitiva $\log_\alpha z$ tiende a α o a $\alpha + 2\pi$ según por qué semiplano nos acerquemos, por lo que no es posible extenderla de forma continua a ninguno de sus puntos.

Consideremos concretamente los logaritmos $\log_{-\pi/2}$ y $\log_{\pi/2}$. El primero está definido salvo en el semieje imaginario negativo y proporciona logaritmos con parte imaginaria entre $-\pi/2$ y $3\pi/2$, mientras que el segundo salvo en el semieje imaginario positivo y proporciona logaritmos con parte imaginaria entre $\pi/2$ y $5\pi/2$.

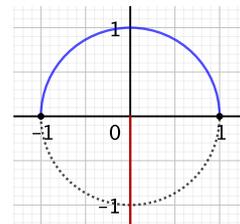
Ahora consideremos la circunferencia $\phi(t) = re^{it}$, para $0 \leq t \leq 2\pi$. Podemos partirla en dos semicircunferencias, ϕ_1 y ϕ_2 , la primera con parámetro $0 \leq t \leq \pi$ y la segunda con parámetro $\pi \leq t \leq 2\pi$. Claramente,

$$\int_{|z|=r} \frac{1}{\zeta} d\zeta = \int_{\phi_1} \frac{1}{\zeta} d\zeta + \int_{\phi_2} \frac{1}{\zeta} d\zeta,$$

pues al expresar la integral como una integral en $[0, 2\pi]$, simplemente tenemos que partirla en suma de una integral en $[0, \pi]$ y otra en $[\pi, 2\pi]$, y éstas son el miembro derecho. La primera integral la podemos calcular con la regla de Barrow usando $\log_{-\pi/2} z$ como primitiva del integrando.

En efecto, la función $\log_{-\pi/2} z$ está definida salvo en el semieje imaginario negativo, luego está definida sobre toda la semicircunferencia ϕ_1 . Por lo tanto:

$$\begin{aligned} \int_{\phi_1} \frac{1}{\zeta} d\zeta &= \log_{-\pi/2}(-r) - \log_{-\pi/2}(r) \\ &= \log r + i\pi - \log r = i\pi. \end{aligned}$$



Similarmente, la integral sobre ϕ_2 se puede calcular usando la primitiva $\log_{\pi/2} z$, y el resultado es

$$\int_{\phi_2} \frac{1}{\zeta} d\zeta = \log_{\pi/2}(-r) - \log_{\pi/2}(r) = \log r + 2\pi i - \log r - i\pi = i\pi.$$

Por consiguiente, la suma de ambas integrales es $2\pi i$, como ya habíamos calculado directamente. El resultado no es 0 porque no podemos usar la misma primitiva para calcular las integrales sobre cada una de las semicircunferencias.

Un argumento alternativo que requeriría más trabajo para plasmarlo con rigor, pero que es ilustrativo, consiste en considerar el logaritmo $\log_0 z$, que está definido salvo en el semieje real positivo y proporciona logaritmos con parte imaginaria entre 0 y 2π . Podemos usarlo para calcular la integral sobre el arco ϕ_ϵ que resulta de restringir ϕ al intervalo $[\epsilon, 2\pi - \epsilon]$, para cualquier $0 < \epsilon < \pi$. El resultado es

$$\int_{\phi_\epsilon} \frac{1}{\zeta} d\zeta = \log_0(re^{i(2\pi-\epsilon)}) - \log_0(re^{i\epsilon}) = \log r + (2\pi - \epsilon)i - \log r - i\epsilon = 2\pi i - 2\epsilon i.$$

Esto significa que, por la regla de Barrow, la integral en “casi toda” la circunferencia tiene que ser “casi” $2\pi i$, por lo que es natural que la integral en toda la circunferencia deba ser $2\pi i$ y no 0. En resumen, el hecho de que la integral sobre la circunferencia de $1/z$ sea no nula se debe en esencia a que si vamos calculando el logaritmo de los puntos de la circunferencia partiendo, por ejemplo, de $\log r$, la parte imaginaria va aumentando desde 0, y cuando volvemos al punto de partida tras haber recorrido la circunferencia, la parte imaginaria no se acerca a 0, sino a 2π . ■

Terminamos con un par de propiedades técnicas. La primera es una condición suficiente para que podamos intercambiar una suma infinita y una integral:

Teorema 7.9 *Sea $\phi : [a, b] \rightarrow \mathbb{C}$ un arco y sea $f_n : \phi^* \rightarrow \mathbb{C}$ una sucesión de funciones continuas. Sea $\{M_n\}_{n=0}^{\infty}$ una sucesión de números reales tal que si $\zeta \in \phi^*$ se cumple que $|f_n(\zeta)| \leq M_n$ y la serie $\sum_{n=0}^{\infty} M_n$ es convergente. Entonces*

$$\int_{\phi} \sum_{n=0}^{\infty} f_n(\zeta) d\zeta = \sum_{n=0}^{\infty} \int_{\phi} f_n(\zeta) d\zeta.$$

DEMOSTRACIÓN: Por el criterio de mayoración de Weierstrass [ITAn 3.7] la serie converge a una función continua en ϕ^* . Además,

$$\left| \sum_{n=0}^N \int_{\phi} f_n(\zeta) d\zeta - \int_{\phi} \sum_{n=0}^{\infty} f_n(\zeta) d\zeta \right| = \left| \int_{\phi} \sum_{n=N+1}^{\infty} f_n(\zeta) d\zeta \right|.$$

Dado $\epsilon > 0$, tomamos N_0 tal que si $N \geq N_0$ entonces $\sum_{n=N+1}^{\infty} M_n < \epsilon/L(\phi)$.

Así

$$\left| \sum_{n=N+1}^{\infty} f_n(\zeta) \right| \leq \sum_{n=N+1}^{\infty} |f_n(\zeta)| \leq \sum_{n=N+1}^{\infty} M_n < \frac{\epsilon}{L(\phi)},$$

y la propiedad c) del teorema 7.6 nos permite concluir que

$$\left| \sum_{n=0}^N \int_{\phi} f_n(\zeta) d\zeta - \int_{\phi} \sum_{n=0}^{\infty} f_n(\zeta) d\zeta \right| < \epsilon$$

para todo $N \geq N_0$, luego la suma de las integrales converge a la integral de la suma. ■

Por último generalizamos el teorema 3.11 sobre derivación de integrales paramétricas al caso de integrales curvilíneas.

Teorema 7.10 *Sea $A \subset \mathbb{C}$ un abierto, sea $\phi : [a, b] \rightarrow \mathbb{C}$ un arco y sea $f : A \times \phi^* \rightarrow \mathbb{C}$ una función continua con derivada compleja respecto a la primera variable $f' : A \times \phi^* \rightarrow \mathbb{C}$, continua. Sea $F : A \rightarrow \mathbb{C}$ la función dada por*

$$F(z) = \int_{\phi} f(z, \zeta) d\zeta.$$

Entonces F es holomorfa en A y su derivada es

$$F'(z) = \int_{\phi} f'(z, \zeta) d\zeta.$$

DEMOSTRACIÓN: Por definición,

$$F(z) = \int_a^b f(z, \phi(t)) \phi'(t) dt.$$

Separando las partes real e imaginaria:

$$\begin{aligned} F(z) &= \int_a^b \operatorname{Re} f(z, \phi(t)) \operatorname{Re} \phi'(t) dt - \int_a^b \operatorname{Im} f(z, \phi(t)) \operatorname{Im} \phi'(t) dt \\ &+ i \int_a^b \operatorname{Re} f(z, \phi(t)) \operatorname{Im} \phi'(t) dt + i \int_a^b \operatorname{Im} f(z, \phi(t)) \operatorname{Re} \phi'(t) dt. \end{aligned}$$

Vamos a aplicar el teorema 3.11 (teniendo en cuenta la nota posterior) a las cuatro integrales. Concretamente, fijado $z = x + yi \in A$, podemos tomar $\delta > 0$ tal que $[x - \delta, x + \delta] \times [y - \delta, y + \delta] \subset A$.

Consideramos la función $f^* :]x - \delta, x + \delta[\times [a, b] \rightarrow \mathbb{R}$ dada por

$$f^*(x, t) = \operatorname{Re} f(x + yi, \phi(t)) \operatorname{Re} \phi'(t),$$

que es continua (recordemos que en la definición de arco hemos exigido que las derivadas de los arcos sean continuas) y existe

$$\frac{\partial f^*}{\partial x} :]x - \delta, x + \delta[\times [a, b] \rightarrow \mathbb{R}$$

dada por

$$\frac{\partial f^*}{\partial x} = \frac{\partial \operatorname{Re} f}{\partial x}(x + yi, \phi(t)) \operatorname{Re} \phi'(t) = \operatorname{Re} f'(x + yi, \phi(t)) \operatorname{Re} \phi'(t),$$

donde hemos usado que

$$f'(z) = \frac{\partial f}{\partial x}.$$

Así, de las hipótesis del teorema se sigue que la función $\partial f^*/\partial x$ es continua en $]x - \delta, x + \delta[\times [a, b]$, luego podemos aplicar el teorema 3.11 a la primera integral, y análogamente se razona con las otras tres. La conclusión es que

$$\begin{aligned} \frac{\partial F}{\partial x} &= \int_a^b \frac{\partial \operatorname{Re} f}{\partial x}(z, \phi(t)) \operatorname{Re} \phi'(t) dt - \int_a^b \frac{\partial \operatorname{Im} f}{\partial x}(z, \phi(t)) \operatorname{Im} \phi'(t) dt \\ &+ i \int_a^b \frac{\partial \operatorname{Re} f}{\partial x}(z, \phi(t)) \operatorname{Im} \phi'(t) dt + i \int_a^b \frac{\partial \operatorname{Im} f}{\partial x}(z, \phi(t)) \operatorname{Re} \phi'(t) dt \\ &= \int_a^b \frac{\partial f}{\partial x}(z, \phi(t)) \phi'(t) dt = \int_{\phi} \frac{\partial f}{\partial x}(z, \zeta) d\zeta. \end{aligned}$$

De aquí se sigue que la función $\partial F/\partial x$ es continua. Como esto es un resultado general que no depende del argumento que demuestra este teorema, vamos a aceptarlo de momento y lo justificaremos al terminar la prueba. El mismo razonamiento nos da la relación

$$\frac{\partial F}{\partial x} = \int_{\phi} \frac{\partial f}{\partial y}(z, \zeta) d\zeta,$$

y en particular la continuidad de la derivada. Así, pues, como F tiene derivadas parciales continuas, el teorema 5.7 nos da que F es diferenciable. Más aún, para cada t fijo, la función $f(z, t)$ cumple las ecuaciones de Cauchy-Riemann, es decir, cumple

$$\frac{\partial f}{\partial y} = i \frac{\partial f}{\partial x},$$

luego de las relaciones que hemos obtenido se sigue inmediatamente que

$$\frac{\partial F}{\partial y} = i \frac{\partial F}{\partial x},$$

luego 7.2 nos da que la función $F(z)$ es holomorfa. ■

Para completar la prueba del teorema anterior nos falta justificar que una función de la forma

$$F(x, y) = \int_a^b f(x, y, t) dt$$

es continua si el integrando $f : A \times [a, b] \rightarrow \mathbb{R}$ es una función continua. En la prueba del teorema anterior, esto se aplica a las partes real e imaginaria de las funciones $\partial F/\partial x$ y $\partial F/\partial y$, para concluir que ambas son funciones continuas.

Fijado un punto (x, y) del dominio de F , podemos considerar un producto $K = [x - \delta, x + \delta] \times [y - \delta, y + \delta] \times [a, b]$ contenido en el dominio de f . Aplicamos a f el teorema de Heine-Cantor³ [ITAn A.14]: dado $\epsilon > 0$, podemos tomar un $\delta > 0$ tal que si $(x, y, t), (x', y', t') \in K$ cumplen $\|(x, y, t) - (x', y', t')\| < \delta$, entonces

$$|f(x, y, t) - f(x', y', t')| < \frac{\epsilon}{b-a}.$$

En particular, para todo $t \in [a, b]$, si $|(x, y) - (x', y')| < \delta$, entonces

$$|F(x, y) - F(x', y')| = \left| \int_a^b (f(x, y, t) - f(x', y', t)) dt \right| \leq \int_a^b |f(x, y, t) - f(x', y', t)| dt \leq \int_a^b \frac{\epsilon}{b-a} dt = \epsilon,$$

lo que prueba la continuidad de F en (x, y) . ■

³En [ITAn] hemos probado el teorema para funciones definidas sobre compactos $K \subset \mathbb{R}^2$, mientras que aquí lo necesitamos para un compacto $K \subset \mathbb{R}^3$, pero el lector puede comprobar sin dificultad que el argumento vale igualmente con los cambios triviales.

Las integrales de Fresnel Veamos ahora cómo se pueden aplicar los conceptos que hemos introducido para dar rigor al cálculo de las integrales de Fresnel presentado en la introducción:

$$\int_0^{+\infty} \cos t^2 dt = \int_0^{+\infty} \operatorname{sen} t^2 dt = \sqrt{\frac{\pi}{8}}.$$

Recordemos que consistía en unir las en una única integral compleja:

$$\int_0^{+\infty} \cos t^2 dt + i \int_0^{+\infty} \operatorname{sen} t^2 dt = \int_0^{+\infty} e^{it^2} dt = \int_0^{+\infty} e^{-(\omega t)^2} dt,$$

donde

$$\omega = \frac{\sqrt{2}}{2} - i \frac{\sqrt{2}}{2},$$

cumple $\omega^2 = i$, y luego hacer el “polémico” cambio de variable $x = \omega t$. En primer lugar observamos que

$$\int_{[0, \omega R]} e^{-\zeta^2} d\zeta = \omega \int_0^R e^{-(\omega t)^2} dt = \omega \int_0^R \cos t^2 dt + \omega i \int_0^R \operatorname{sen} t^2 dt,$$

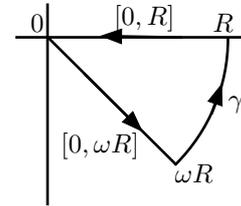
pues el segmento $[0, \omega R]$ puede parametrizarse como $\phi(t) = \omega t$, con $0 \leq t \leq R$. Igualmente

$$\int_{[0, R]} e^{-t^2} dt = \int_0^R e^{-t^2} dt$$

y por último consideramos el arco $\gamma : [-\pi/2, 0] \rightarrow \mathbb{C}$ dado por $\gamma(t) = Re^{it}$.

Vamos a admitir que la función holomorfa e^{-z^2} admite una primitiva en \mathbb{C} . Esto lo demostraremos más adelante, pero, aceptándolo de momento, la regla de Barrow nos da que

$$\int_{[0, \omega R]} e^{-\zeta^2} d\zeta + \int_{\gamma} e^{-\zeta^2} d\zeta - \int_{[0, R]} e^{-\zeta^2} d\zeta = 0,$$



pues si F es una primitiva, la suma de estas integrales es

$$F(\omega R) - F(0) + F(R) - F(\omega R) + F(0) - F(R) = 0.$$

Explícitamente,

$$\int_0^R \cos t^2 dt + i \int_0^R \operatorname{sen} t^2 dt = \int_0^R e^{-t^2} dt - \int_{-\pi/4}^0 e^{-R^2 e^{2it}} i R e^{it} dt.$$

Si probamos que la última integral tiende a 0 cuando R tiende a $+\infty$, podremos concluir que

$$\int_0^{+\infty} \cos t^2 dt + i \int_0^{+\infty} \operatorname{sen} t^2 dt = \omega^{-1} \int_0^{+\infty} e^{-t^2} dt = \left(\frac{\sqrt{2}}{2} + i \frac{\sqrt{2}}{2} \right) \frac{\sqrt{\pi}}{2},$$

de donde se sigue inmediatamente el valor anunciado de las integrales de Fresnel. Ahora bien,

$$\left| \int_{-\pi/4}^0 e^{-R^2 e^{2it}} iR e^{it} dt \right| \leq R \int_{-\pi/4}^0 |e^{-R^2 e^{2it}}| dt \leq R \int_{-\pi/4}^0 e^{-R^2 \cos 2t} dt.$$

El cambio de variable $t = t' - \pi/4$ sustituye el coseno por un seno y, dado que la función $(\operatorname{sen} x)/x$ es decreciente en $[0, \pi/2]$, resulta que $\operatorname{sen} 2t \geq 4t/\pi$, para $0 \leq t \leq \pi/4$, y el módulo de la integral queda acotado por

$$R \int_0^{\pi/4} e^{-4R^2 t/\pi} dt = -\frac{\pi}{4R} \left[e^{-4R^2 t/\pi} \right]_0^{\pi/4} = \frac{\pi(1 - e^{-R^2})}{4R}.$$

El último término tiende claramente a 0.

Así pues, lo único que nos falta para completar el cálculo de las integrales de Fresnel es demostrar que la función e^{-z^2} tiene primitiva en \mathbb{C} . Equivalentemente (salvo este punto pendiente), hemos demostrado que

$$\int_{-\infty}^{+\infty} \cos t^2 dt = \int_{-\infty}^{+\infty} \operatorname{sen} t^2 dt = \sqrt{\frac{\pi}{2}},$$

o también, según (3.2), que

$$\int_0^{+\infty} \frac{\cos u}{\sqrt{u}} du = \int_0^{+\infty} \frac{\operatorname{sen} u}{\sqrt{u}} du = \sqrt{\frac{\pi}{2}}.$$

La integral de Dirichlet Por último aplicamos la misma técnica para dar un cálculo alternativo de la integral de Dirichlet:

$$\int_0^{+\infty} \frac{\operatorname{sen} x}{x} dx = \frac{\pi}{2}.$$

Para ello consideramos la función $f(z) = e^{iz}/z$, que es holomorfa en $\mathbb{C} \setminus \{0\}$. Fijados $0 < r < R$, observamos que

$$\begin{aligned} \int_{[r,R]} \frac{e^{i\zeta}}{\zeta} d\zeta &= \int_r^R \frac{e^{ix}}{x} dx = \int_r^R \frac{\cos x}{x} dx + i \int_r^R \frac{\operatorname{sen} x}{x} dx, \\ \int_{[-R,-r]} \frac{e^{i\zeta}}{\zeta} d\zeta &= \int_{-r}^{-R} \frac{e^{ix}}{x} dx = - \int_r^R \frac{\cos x}{x} dx + i \int_r^R \frac{\operatorname{sen} x}{x} dx, \end{aligned}$$

con lo que

$$\int_{[-R,-r]} \frac{e^{i\zeta}}{\zeta} d\zeta + \int_{[r,R]} \frac{e^{i\zeta}}{\zeta} d\zeta = 2i \int_r^R \frac{\operatorname{sen} x}{x} dx.$$

Además consideramos los arcos $\phi_r : [0, \pi] \rightarrow \mathbb{C}$ dados por $\phi_r(t) = r e^{it}$. No es cierto que $f(z)$ tenga primitiva en $\mathbb{C} \setminus \{0\}$, pero veremos que la tiene en el

abierto que resulta de quitarle a $\mathbb{C} \setminus \{0\}$ el semieje imaginario negativo. Esto basta para asegurar que, en virtud de la regla de Barrow, se cumple que

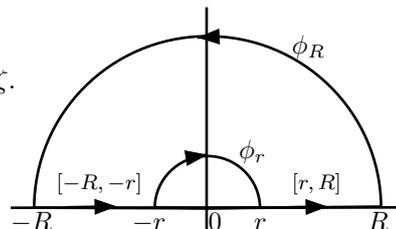
$$\int_{[-R, -r]} f(\zeta) d\zeta - \int_{\phi_r} f(\zeta) d\zeta + \int_{[r, R]} f(\zeta) d\zeta + \int_{\phi_R} f(\zeta) d\zeta = 0$$

o, equivalentemente,

$$2i \int_r^R \frac{\operatorname{sen} x}{x} dx = \int_{\phi_r} f(\zeta) d\zeta - \int_{\phi_R} f(\zeta) d\zeta.$$

Ahora calculamos:

$$\int_{\phi_r} \frac{e^{i\zeta}}{\zeta} d\zeta = i \int_0^\pi e^{ire^{it}} dt.$$



Como e^{iz} es continua en 0, dado $\epsilon > 0$ existe un $\delta > 0$ tal que si $|z| < \delta$ entonces $|e^{iz} - 1| < \epsilon/\pi$. Por lo tanto, si $0 < r < \epsilon$, se cumple que

$$\left| \int_{\phi_r} \frac{e^{i\zeta}}{\zeta} d\zeta - i\pi \right| = \left| i \int_0^\pi e^{ire^{it}} dt - i \int_0^\pi dt \right| \leq \int_0^\pi |e^{ire^{it}} - 1| dt \leq \epsilon,$$

luego

$$\lim_{r \rightarrow 0} \int_{\phi_r} \frac{e^{i\zeta}}{\zeta} d\zeta = i\pi,$$

luego

$$2i \int_0^R \frac{\operatorname{sen} x}{x} dx = i\pi - \int_{\phi_R} f(\zeta) d\zeta.$$

Ahora basta probar que la última integral tiende a 0 cuando $R \rightarrow +\infty$, pues entonces podremos concluir que

$$2i \int_0^{+\infty} \frac{\operatorname{sen} x}{x} dx = i\pi.$$

En efecto:

$$\begin{aligned} \left| \int_{\phi_R} \frac{e^{i\zeta}}{\zeta} d\zeta \right| &= \left| \int_0^\pi e^{iRe^{it}} dt \right| \leq \int_0^\pi |e^{iR \cos t - R \operatorname{sen} t}| dt = \int_0^\pi e^{-R \operatorname{sen} t} dt \\ &= \int_0^{\pi/2} e^{-R \operatorname{sen} t} dt + \int_{\pi/2}^\pi e^{-R \operatorname{sen} t} dt = 2 \int_0^{\pi/2} e^{-R \operatorname{sen} t} dt, \end{aligned}$$

porque la segunda integral se transforma en la primera con el cambio $t = \pi - t'$.

Ahora usamos que si $0 \leq t \leq \pi/2$, entonces, como la función $\operatorname{sen} t/t$ es decreciente en este intervalo,⁴

$$\frac{\operatorname{sen} t}{t} \geq \frac{\operatorname{sen}(\pi/2)}{\pi/2} = \frac{2}{\pi},$$

⁴Derivándola, basta ver que $t \cos t - \operatorname{sen} t \leq 0$, para lo cual, volviendo a derivar, basta ver que $\cos t - t \operatorname{sen} t - \cos t = -t \operatorname{sen} t \leq 0$, lo cual es obvio.

luego

$$\left| \int_{\phi_R} \frac{e^{i\zeta}}{\zeta} d\zeta \right| \leq 2 \int_0^{\pi/2} e^{-2Rt/\pi} dt = \pi \frac{1 - e^{-R}}{R} < \frac{\pi}{R},$$

y la última expresión tiende a 0 cuando $R \rightarrow +\infty$. ■

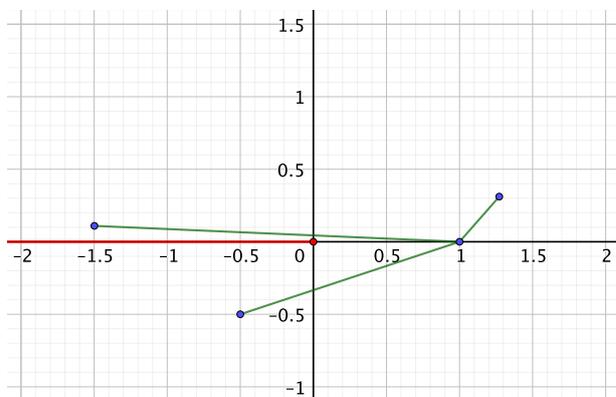
7.3 Propiedades de las funciones holomorfas

En la sección anterior hemos visto que no toda función holomorfa definida en un abierto tiene primitiva en él. Esto puede parecer un comportamiento “peor” que el de las funciones de variable real, pues sabemos que toda función continua definida en un intervalo cerrado tiene primitiva. Sin embargo, lo que sucede es que la existencia de primitiva está también garantizada en el caso complejo bajo una hipótesis sobre el dominio de la función que se cumple trivialmente en el caso real, en el que sólo consideramos funciones definidas sobre intervalos.

Definición 7.11 Un conjunto $A \subset \mathbb{C}$ es *estrellado* con centro en $z_0 \in A$ si, para todo $z \in A$, el segmento $[z_0, z]^* \subset A$.

Por ejemplo, el abierto $A = \mathbb{C} \setminus \{0\}$ no es un conjunto estrellado, pues, cualquiera que sea el punto $z_0 \in \mathbb{C} \setminus \{0\}$ que elijamos, el segmento $[z_0, -z_0]^*$ no está contenido en A , ya que pasa por 0.

En cambio, el abierto \mathbb{C}_- que resulta de quitarle a $\mathbb{C} \setminus \{0\}$ el semieje real negativo sí que es estrellado, pues cualquiera de sus puntos puede unirse con $z_0 = 1$ mediante un segmento contenido en \mathbb{C}_- .



Vamos a probar que toda función holomorfa en un abierto estrellado tiene primitiva. Un abierto A es *convexo* si cuando $a, b \in A$, el segmento $[a, b]^*$ está contenido en A .

Obviamente todo abierto convexo es estrellado tomando como centro cualquiera de sus puntos. Entre los conjuntos convexos que más nos van a interesar están los discos abiertos:

$$D(z_0, r) = \{z \in \mathbb{C} \mid |z - z_0| < r\}$$

y los cerrados:

$$\overline{D(z_0, r)} = \{z \in \mathbb{C} \mid |z - z_0| \leq r\}.$$

En efecto, si $a, b \in D(z_0, r)$, un elemento $z \in [a, b]^*$ es de la forma

$$z = (1-t)a + tb, \quad 0 \leq t \leq 1.$$

Entonces

$$\begin{aligned} |z - z_0| &= |(1-t)a + tb - ((1-t)z_0 + tz_0)| \leq \\ &(1-t)|a - z_0| + t|b - z_0| < (1-t)r + tr = r, \end{aligned}$$

luego $z \in D(z_0, r)$, e igualmente se razona con el disco cerrado.

Ahora necesitamos algunas consideraciones sobre triángulos. Vamos a llamar *triángulo* a toda terna $T = (a, b, c)$ de números complejos (sin excluir la posibilidad de que coincidan o de que estén alineados). Los lados del triángulo serán los arcos $[a, b]$, $[b, c]$ y $[c, a]$. La longitud de un triángulo la definimos como la suma de las longitudes de sus lados:

$$L(T) = |b - a| + |c - b| + |a - c|.$$

El *contenido* de un triángulo $T = (a, b, c)$ es

$$[T] = \{\alpha a + \beta b + \gamma c \mid \alpha, \beta, \gamma \in [0, 1], \alpha + \beta + \gamma = 1\}.$$

Geoméricamente $[T]$ no es sino el conjunto de los puntos rodeados por los lados del triángulo. Esto es intuitivamente inmediato a partir del hecho siguiente:

Un número complejo z está en $[T]$ si y sólo si está en $[a, x]^$, para cierto $x \in [b, c]^*$.*

En efecto, si

$$z = \alpha a + \beta b + \gamma c \in [T],$$

o bien $\alpha = 1$, en cuyo caso $z = a$, o bien $\alpha \neq 1$, en cuyo caso

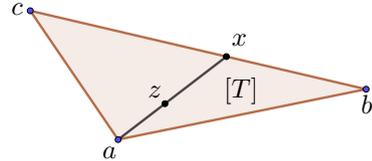
$$z = \alpha a + (1-\alpha) \left(\frac{\beta}{\beta+\gamma} b + \frac{\gamma}{\beta+\gamma} c \right) \quad \text{y} \quad x = \frac{\beta}{\beta+\gamma} b + \frac{\gamma}{\beta+\gamma} c \in [b, c]^*.$$

En cualquier caso $z \in [a, x]^*$ para un cierto $x \in [b, c]^*$. Recíprocamente, si z cumple esto, entonces $x = (1-\gamma)b + \gamma c$, con $0 \leq \gamma \leq 1$ y $z = (1-\beta)a + \beta x$, con $0 \leq \beta \leq 1$. Así $z = (1-\beta)a + \beta(1-\gamma)b + \beta\gamma c$ y los coeficientes suman 1, luego $z \in [T]$. ■

Es claro entonces que si A es convexo y $a, b, c \in A$, entonces $[T] \subset A$.

Si $f : A \rightarrow \mathbb{C}$ es una función continua y T es un triángulo cuyos lados están contenidos en A , podemos definir

$$\int_T f(\zeta) d\zeta = \int_{[a,b]} f(\zeta) d\zeta + \int_{[b,c]} f(\zeta) d\zeta + \int_{[c,a]} f(\zeta) d\zeta.$$



Si f tiene primitiva F en A , por la regla de Barrow la integral anterior es nula, pues

$$\int_T f(\zeta) d\zeta = F(b) - F(a) + F(c) - F(b) + F(a) - F(c) = 0.$$

Ahora vamos a demostrar un recíproco:

Teorema 7.12 *Sea $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ una función continua definida sobre un abierto estrellado de centro z_0 tal que para todo triángulo T tal que $[T] \subset A$ se cumple que*

$$\int_T f(\zeta) d\zeta = 0.$$

Entonces la función $F : A \rightarrow \mathbb{C}$ dada por

$$F(z) = \int_{[z_0, z]} f(\zeta) d\zeta$$

es holomorfa en A y $F' = f$.

DEMOSTRACIÓN: Fijemos $z_1 \in A$ y vamos a probar que F es derivable en z_1 . Como A es abierto, existe un $r > 0$ tal que $D(z_1, r) \subset A$. Sea $\epsilon > 0$. Como f es continua en z_1 , existe un δ tal que $0 < \delta < r$ y si $|z - z_1| < \delta$, entonces $|f(z) - f(z_1)| < \epsilon$.

Sea $z \neq z_1$ tal que $|z - z_1| < \delta$. Por definición de F :

$$\left| \frac{F(z) - F(z_1)}{z - z_1} - f(z_1) \right| = \left| \frac{1}{z - z_1} \left(\int_{[z_0, z]} f(\zeta) d\zeta - \int_{[z_0, z_1]} f(\zeta) d\zeta \right) - f(z_1) \right|.$$

Ahora consideramos el triángulo $T = (z_0, z, z_1)$ y observamos que $[T] \subset A$, pues, como el disco $D(z_1, \delta)$ es convexo, tenemos que $[z, z_1]^* \subset D(z_1, \delta) \subset A$, y todo punto de $[T]$ está en un segmento de extremos z_0 y un punto de $[z, z_1]^*$, luego está en A por definición de abierto estrellado. Por hipótesis

$$\int_{[z_0, z]} f(\zeta) d\zeta + \int_{[z, z_1]} f(\zeta) d\zeta + \int_{[z_1, z_0]} f(\zeta) d\zeta = 0,$$

lo cual equivale a que

$$\int_{[z_0, z]} f(\zeta) d\zeta - \int_{[z_0, z_1]} f(\zeta) d\zeta = \int_{[z_1, z]} f(\zeta) d\zeta,$$

luego

$$\begin{aligned} \left| \frac{F(z) - F(z_1)}{z - z_1} - f(z_1) \right| &= \left| \frac{1}{z - z_1} \int_{[z_1, z]} f(\zeta) d\zeta - \frac{1}{z - z_1} \int_{[z_1, z]} f(z_1) d\zeta \right| \\ &\leq \frac{1}{|z - z_1|} \int_{[z_1, z]} |f(\zeta) - f(z_1)| d\zeta. \end{aligned}$$

Pero $[z_1, z]^* \subset D(z_1, \delta)$, luego el integrando está acotado por ϵ , luego queda

$$\left| \frac{F(z) - F(z_1)}{z - z_1} - f(z_1) \right| \leq \epsilon, \quad \text{si } |z - z_1| < \delta.$$

Por lo tanto F es derivable en z_1 y $F'(z_1) = f(z_1)$. ■

Observemos que el teorema anterior es el análogo al hecho de que, dada una función continua $f :]a, b[\rightarrow \mathbb{R}$ y $x_0 \in]a, b[$, la función

$$F(x) = \int_{x_0}^x f(t) dt$$

es una primitiva de f en $]a, b[$, pero en el caso complejo tenemos que asegurarnos de que al movernos desde z_0 hasta z no nos salimos del dominio de definición de f , cosa que en el caso real no puede suceder.

Demostraremos que la condición de que las integrales sobre triángulos sean nulas se cumple siempre que el integrando es una función holomorfa. Esto es un resultado nada trivial y que está en la base de todas las propiedades “sorprendentes” de las funciones holomorfas. Lo enunciaremos a continuación y lo demostraremos en la sección siguiente:

Teorema 7.13 (Teorema de Cauchy para triángulos) *Sea A un abierto en \mathbb{C} , sea $p \in A$ y sea $f : A \rightarrow \mathbb{C}$ una función continua en A y holomorfa en $A \setminus \{p\}$. Entonces, para todo triángulo T tal que $[T] \subset A$ se cumple que*

$$\int_T f(\zeta) d\zeta = 0.$$

Observemos que no hemos exigido realmente que el integrando sea una función holomorfa en A , sino que permitimos que haya a lo sumo un punto excepcional. Esto es un tecnicismo al que le sacaremos partido para demostrar, entre otras cosas, que no puede existir tal punto excepcional: veremos, en efecto (véase la observación tras el teorema 7.16), que si una función es derivable salvo a lo sumo en un punto p , pero es continua en p , entonces también es derivable en p .

Esto es falso para funciones de variable real. Por ejemplo, $f(x) = |x|$ es continua en \mathbb{R} y derivable en $\mathbb{R} \setminus \{0\}$, pero no en 0.

De momento, aceptando el teorema de Cauchy, tenemos probado lo siguiente:

Teorema 7.14 *Sea $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ una función continua en un abierto estrellado A y holomorfa en $A \setminus \{p\}$, para cierto punto $p \in A$. Entonces existe una función holomorfa $F : A \rightarrow \mathbb{C}$ tal que $F' = f$.*

En particular, la función e^{-z^2} tiene primitiva en \mathbb{C} y la función e^{iz}/z tiene primitiva en el abierto que resulta de quitarle a $\mathbb{C} \setminus \{0\}$ el semieje imaginario negativo, lo que completa el cálculo de las integrales de Fresnel y de Dirichlet que hemos hecho en la sección precedente (salvo que tenemos pendiente la demostración del teorema de Cauchy para triángulos).

A su vez, el teorema anterior, junto con la regla de Barrow 7.8, nos permite generalizar el teorema de Cauchy:⁵

Teorema 7.15 (Teorema de Cauchy) *Sea $f : A \rightarrow \mathbb{C}$ una función holomorfa en un abierto estrellado A . Entonces, para todo arco cerrado ϕ contenido en A , se cumple que*

$$\int_{\phi} f(\zeta) d\zeta = 0.$$

Más en general, como hemos usado en el cálculo de las integrales de Fresnel y de Dirichlet, si tenemos una sucesión finita de arcos ϕ_1, \dots, ϕ_n de modo que cada uno empieza donde termina el anterior y el último termina donde empieza el primero, de modo que todos ellos forman un circuito cerrado, en las condiciones del teorema anterior, la suma de las integrales sobre ellos es 0.

Ahora estamos en condiciones de probar un hecho sorprendente para cualquiera familiarizado con el cálculo diferencial de variable real: si una función compleja es derivable, de hecho es infinitamente derivable:

Teorema 7.16 (Fórmulas de Cauchy para circunferencias) *Sea A un abierto en \mathbb{C} y $f : A \rightarrow \mathbb{C}$ una función holomorfa. Entonces f es infinitamente derivable y, si $z_0 \in A$ y $r > 0$ es tal que $\overline{D(z_0, r)} \subset A$, entonces, si $|z - z_0| < r$ y n es un número natural,*

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_{|\zeta - z_0| = r} \frac{f(\zeta)}{(\zeta - z)^{n+1}} d\zeta.$$

DEMOSTRACIÓN: Comenzamos probando la fórmula para $n = 0$. Como A es abierto, dado $z_0 \in A$, existe un $r' > 0$ tal que $D(z_0, r') \subset A$. Probaremos la fórmula para $0 < r < r'$, lo cual es suficiente para todas las aplicaciones, aunque es fácil ver que, fijado r en las condiciones del enunciado, siempre podemos tomar $r' > r$ que cumpla estas mismas condiciones, luego en realidad no estamos perdiendo generalidad.

Para cada $z \in D(z_0, r')$, consideremos la función en $D(z_0, r')$ dada por

$$g(\zeta) = \begin{cases} \frac{f(\zeta) - f(z)}{\zeta - z} & \text{si } \zeta \neq z \\ f'(z) & \text{si } \zeta = z \end{cases}$$

Claramente es continua en $D(z_0, r')$ y derivable salvo quizá en z . Por 7.14 tiene primitiva en $D(z_0, r')$ y, por el teorema 7.8,

$$\begin{aligned} 0 &= \int_{|\zeta - z_0| = r} g(\zeta) d\zeta = \int_{|\zeta - z_0| = r} \frac{f(\zeta) - f(z)}{\zeta - z} d\zeta \\ &= \int_{|\zeta - z_0| = r} \frac{f(\zeta)}{\zeta - z} d\zeta - f(z) \int_{|\zeta - z_0| = r} \frac{1}{\zeta - z} d\zeta. \end{aligned}$$

⁵En realidad el teorema de Cauchy se puede generalizar a funciones holomorfas definidas sobre abiertos más generales que los convexos (los llamados abiertos "simplemente conexos", como por ejemplo \mathbb{C}_-), pero no vamos a verlo aquí.

Pero la última integral vale $2\pi i$ por (7.3), lo que nos da la fórmula buscada. Admitiendo la validez de la fórmula para n , el teorema 7.10 implica que $f^{(n)}$ es derivable y su derivada es la dada por la fórmula para $n + 1$. ■

En particular, ahora vemos que si $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ es continua en A y holomorfa en $A \setminus \{p\}$, podemos tomar un disco A_0 centrado en p contenido en A , que es un abierto convexo, y la restricción de f a A_0 satisface las hipótesis del teorema 7.14, luego tiene una primitiva holomorfa F , luego $f = F'$, pero acabamos de probar que las derivadas de las funciones holomorfas son holomorfas, luego f es holomorfa en todo A_0 , y en particular es derivable en p .

Veamos una última consecuencia inmediata de los teoremas de Cauchy.

Teorema 7.17 (Teorema de Morera) *Si $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ es una función continua en un abierto A con la propiedad de que para todo triángulo T tal que $[T] \subset A$ se cumple que $\int_T f(\zeta) d\zeta = 0$. Entonces f es holomorfa en A .*

DEMOSTRACIÓN: Sea $z_0 \in A$ y $r > 0$ tal que $D(z_0, r) \subset A$. Entonces $D(z_0, r)$ es un abierto convexo, luego si un triángulo T tiene sus vértices en $D(z_0, r)$, se cumple que $[T] \subset D(z_0, r) \subset A$, luego por hipótesis la integral de f sobre T es nula. Por el teorema 7.12 existe una función holomorfa $F : D(z_0, r) \rightarrow \mathbb{C}$ tal que $F' = f$. Por el teorema anterior f es holomorfa en $D(z_0, r)$, luego en A . ■

De aquí podemos deducir a su vez la versión compleja del teorema 1.21 (compárese con el ejemplo de Weierstrass de la página 40 y la nota posterior):

Teorema 7.18 *Sea $A \subset \mathbb{C}$ un abierto y sea $f_n : A \rightarrow \mathbb{C}$ una sucesión de funciones holomorfas en A . Sea $\{M_n\}_{n=0}^{\infty}$ una sucesión de números reales tal que $|f_n(z)| \leq M_n$ para todo $z \in A$ y de modo que la serie $\sum_{n=0}^{\infty} M_n$ sea convergente. Entonces la serie $\sum_{n=0}^{\infty} f_n(z)$ converge en A a una función holomorfa f tal que*

$$f'(z) = \sum_{n=0}^{\infty} f'_n(z).$$

DEMOSTRACIÓN: Por el criterio de mayoración de Weierstrass [ITAn 3.7] sabemos que la serie de funciones converge a una función continua en A :

$$f(z) = \sum_{n=0}^{\infty} f_n(z).$$

Para probar que f es derivable en un punto $z_0 \in A$, podemos restringirla a un disco $A_0 = D(z_0, r) \subset A$ y probar que si T es un triángulo tal que $[T] \subset A_0$, la integral de f sobre T es nula. Como las funciones f_n son holomorfas en A_0 , el teorema de Cauchy nos da que

$$\int_T f_n(\zeta) d\zeta = 0.$$

Dado $\epsilon > 0$, podemos tomar $N > 0$ tal que

$$\sum_{n=N+1}^{\infty} M_n < \frac{\epsilon}{L(T)},$$

(podemos suponer que el triángulo no tiene longitud nula, pues en tal caso la integral sobre T es trivialmente nula). Así

$$\left| \int_T f(\zeta) d\zeta \right| = \left| \sum_{n=0}^N \int_T f_n(\zeta) d\zeta + \int_T \sum_{n=N+1}^{\infty} f_n(\zeta) d\zeta \right| \leq L(T) \frac{\epsilon}{L(T)} = \epsilon,$$

donde hemos usado la propiedad c) del teorema 7.6, pues el integrando de la última integral cumple

$$\left| \sum_{n=N+1}^{\infty} f_n(\zeta) \right| \leq \sum_{n=N+1}^{\infty} |f_n(\zeta)| \leq \sum_{n=N+1}^{\infty} M_n < \frac{\epsilon}{L(T)}.$$

Como esto vale para todo $\epsilon > 0$, la integral es nula y f es holomorfa.

Por otra parte, si tomamos $r > 0$ suficientemente pequeño como para que $D(z_0, r) \subset A$, la fórmula de Cauchy para la derivada nos da que

$$f'_n(z_0) = \frac{1}{2\pi i} \int_{|\zeta-z_0|=r} \frac{f_n(\zeta)}{(\zeta-z_0)^2} d\zeta,$$

$$f'(z_0) = \frac{1}{2\pi i} \int_{|\zeta-z_0|=r} \frac{f(\zeta)}{(\zeta-z_0)^2} d\zeta.$$

Por lo tanto, si tomamos N_0 tal que si $N > N_0$

$$\sum_{n=N+1}^{\infty} M_n < r\epsilon,$$

tenemos que

$$\left| \sum_{n=0}^N f'_n(z_0) - f'(z_0) \right| = \frac{1}{2\pi} \left| \int_{|\zeta-z_0|=r} \frac{1}{(\zeta-z_0)^2} \sum_{n=N+1}^{\infty} f_n(\zeta) d\zeta \right| \leq \frac{1}{r} r\epsilon = \epsilon,$$

y esto prueba que $f'(z_0)$ es la suma de la serie de las derivadas. ■

Como consecuencia podemos probar que muchas funciones que conocemos son holomorfas:

Teorema 7.19 *La función $f(z) = \sum_{n=0}^{\infty} a_n z^n$ definida por una serie de potencias de radio de convergencia $R > 0$ es holomorfa en su disco de convergencia D_R (entendiendo que si $R = \infty$ entonces f es holomorfa en \mathbb{C}) y*

$$f'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

En efecto, en [ITAn 3.8] probamos que la serie es continua usando el criterio de mayoración de Weierstrass. El teorema anterior tiene las mismas hipótesis (teniendo en cuenta que las funciones $a_n z^n$ son holomorfas, porque son polinomios) y nos da la conclusión. Similarmente, en [ITAn 8.19] demostramos la continuidad de las series de Dirichlet usando también el criterio de mayoración, luego ahora podemos concluir:

Teorema 7.20 *Toda serie de Dirichlet converge a una función holomorfa en su semiplano de convergencia absoluta $\operatorname{Re} z > \sigma_a$, y la función*

$$f(z) = \sum_{n=1}^{\infty} \frac{a_n}{n^z}$$

Además

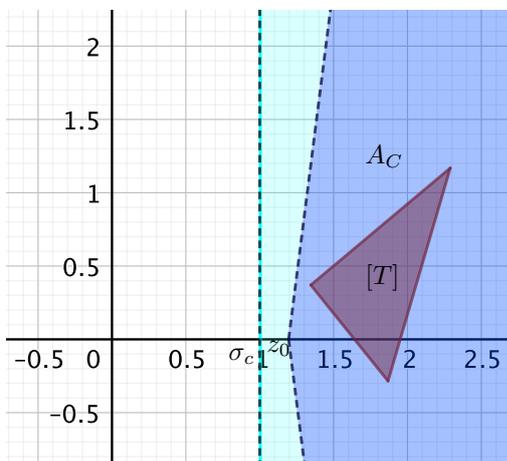
$$f'(z) = \sum_{n=1}^{\infty} \frac{a_n \log n}{n^z}.$$

En particular, la función dseta de Riemann:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

es holomorfa en el semiplano $\operatorname{Re} s > 1$.

Nota En realidad el teorema anterior es válido en todo el semiplano de convergencia de la serie, $\operatorname{Re} z > \sigma_c$, aunque para probarlo tenemos que refinar el argumento.



Para ello observamos que en el teorema [ITAn 8.29] se prueba que, si la serie converge en un punto z_0 , entonces, llamando

$$A_C = \{x + yi \in \mathbb{C} \mid x \geq x_0, |x + yi - z_0| < C|x - x_0|\},$$

para todo $C > 1$ y todo $\epsilon > 0$, existe un número natural N tal que si $m \geq N$ y $z \in A_C$, entonces⁶

$$\left| \sum_{n=m}^{\infty} \frac{a_n}{z^n} \right| < \epsilon.$$

Ahora sólo tenemos que observar que si T es un triángulo tal que $[T]$ está contenido en el semiplano de convergencia de la serie, siempre es posible encontrar un z_0 y un $C > 1$ tales que $[T] \subset A_C$. De este modo,

$$\left| \int_T f(\zeta) d\zeta \right| \leq \left| \int_T \sum_{n=1}^N \frac{a_n}{n^\zeta} d\zeta \right| + \int_T \left| \sum_{n=N+1}^{\infty} \frac{a_n}{n^\zeta} \right| d\zeta \leq L(T)\epsilon,$$

donde hemos usado que la integral de la suma finita es nula porque el integrando es una función holomorfa. Como esto vale para todo $\epsilon > 0$, concluimos que la primera integral es nula, luego f es holomorfa por el teorema de Morera. ■

Similarmente, ahora sabemos que, bajo las hipótesis del teorema [ITAn 8.9] aplicado a una sucesión de funciones holomorfas, podemos concluir que el producto infinito define una función holomorfa. En particular, en virtud del teorema [ITAn 8.15], que se prueba usando [ITAn 8.9], ahora podemos afirmar que el producto infinito

$$P(z) = \prod_{k=1}^{\infty} \left(1 + \frac{z}{k}\right) e^{-z/k}$$

define una función entera, es decir, una función holomorfa en todo el plano complejo, que se anula exactamente en los enteros negativos. A su vez, la función factorial, que, según [ITAn 8.16], puede definirse como

$$\Pi(z) = \frac{1}{e^{\gamma z} P(z)},$$

es una función holomorfa en todo el plano complejo excepto en los enteros negativos, donde tiende a ∞ .

A continuación vamos a demostrar que toda función holomorfa se puede expresar como serie de potencias en cada disco abierto en el que esté definida. Por ejemplo, consideremos la serie geométrica de razón $-x^2$:

$$\sum_{n=0}^{\infty} (-1)^n x^{2n} = \frac{1}{1+x^2}.$$

El miembro derecho es una función derivable en todo \mathbb{R} , pero la serie sólo converge en el intervalo $] -1, 1[$. La razón por la cual la serie no converge en intervalos mayores se entiende cuando consideramos ambos miembros como funciones de variable compleja. Entonces la serie converge en el disco $D(0, 1)$ y no puede converger en un disco mayor porque el miembro derecho no está definido

⁶En la prueba del teorema [ITAn 8,29] se prueba para la suma desde $n = m$ hasta $n = k$, pero, sabiendo que la serie converge, podemos tomar el límite cuando k tiende a ∞ .

en $\pm i$. Es la discontinuidad en $\pm i$ la que “detiene” la convergencia de la serie. El teorema siguiente afirma que esta situación es general: el radio de convergencia de una serie de potencias siempre es el mayor posible:

Teorema 7.21 Sea $f : D(z_0, R) \rightarrow \mathbb{C}$ una función holomorfa. Entonces

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n,$$

para todo $z \in D(z_0, R)$. Además el desarrollo es único.

DEMOSTRACIÓN: Sea $z \in D(z_0, R)$ y sea ρ tal que $|z - z_0| < \rho < R$. Tenemos $z \in D(z_0, \rho) \subset \overline{D}(z_0, \rho) \subset D(z_0, R)$. Podemos aplicar la fórmula de Cauchy, que nos da

$$\begin{aligned} f(z) &= \frac{1}{2\pi i} \int_{|\zeta - z_0| = \rho} \frac{f(\zeta)}{\zeta - z} d\zeta = \frac{1}{2\pi i} \int_{|\zeta - z_0| = \rho} \frac{f(\zeta)}{(\zeta - z_0) - (z - z_0)} d\zeta \\ &= \frac{1}{2\pi i} \int_{|\zeta - z_0| = \rho} \frac{f(\zeta)}{\zeta - z_0} \frac{1}{1 - \frac{z - z_0}{\zeta - z_0}} d\zeta = \frac{1}{2\pi i} \int_{|\zeta - z_0| = \rho} \frac{f(\zeta)}{\zeta - z_0} \sum_{n=0}^{\infty} \left(\frac{z - z_0}{\zeta - z_0} \right)^n d\zeta \\ &= \frac{1}{2\pi i} \int_{|\zeta - z_0| = \rho} \sum_{n=0}^{\infty} \frac{f(\zeta)}{\zeta - z_0} \left(\frac{z - z_0}{\zeta - z_0} \right)^n d\zeta. \end{aligned}$$

Vamos a probar que podemos intercambiar la suma y la integral. Aceptándolo de momento, podemos continuar:

$$\begin{aligned} f(z) &= \frac{1}{2\pi i} \sum_{n=0}^{\infty} \int_{|\zeta - z_0| = \rho} \frac{f(\zeta)}{\zeta - z_0} \left(\frac{z - z_0}{\zeta - z_0} \right)^n d\zeta \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{2\pi i} \int_{|\zeta - z_0| = \rho} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta \right) (z - z_0)^n \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!} (z - z_0)^n, \end{aligned}$$

por las fórmulas de Cauchy para las derivadas.

Para probar el intercambio aplicamos el teorema 7.9. Para ello basta observar que la circunferencia de centro z_0 y radio ρ es compacta, luego existe un $M > 0$ tal que

$$\left| \frac{f(\zeta)}{\zeta - z_0} \right| < M$$

para todo ζ en la circunferencia.

Por consiguiente,

$$\left| \frac{f(\zeta)}{\zeta - z_0} \left(\frac{z - z_0}{\zeta - z_0} \right)^n \right| \leq M \left(\frac{|z - z_0|}{\rho} \right)^n = M_n,$$

y la serie $\sum_{n=0}^{\infty} M_n$ es convergente.

Para probar la unicidad del desarrollo suponemos que dos series de potencias definen la misma función holomorfa en $D(z_0, R)$:

$$\sum_{n=0}^{\infty} a_n(z - z_0)^n = \sum_{n=0}^{\infty} b_n(z - z_0)^n.$$

Tenemos que probar que $a_n = b_n$ para todo n . En caso contrario, sea n_0 el menor índice para el que $a_{n_0} \neq b_{n_0}$. Entonces, para todo $z \in D(z_0, R)$, tenemos que

$$\sum_{n=n_0}^{\infty} a_n(z - z_0)^n = \sum_{n=n_0}^{\infty} b_n(z - z_0)^n,$$

luego

$$(z - z_0)^{n_0} \sum_{n=n_0}^{\infty} a_n(z - z_0)^{n-n_0} = (z - z_0)^{n_0} \sum_{n=n_0}^{\infty} b_n(z - z_0)^{n-n_0},$$

luego, para todo $z \in D(z_0, R)$, $z \neq z_0$, tenemos que

$$\sum_{n=n_0}^{\infty} a_n(z - z_0)^{n-n_0} = \sum_{n=n_0}^{\infty} b_n(z - z_0)^{n-n_0},$$

pero ambas series definen funciones holomorfas en $D(z_0, R)$, que en particular serán continua y, como en $D(z_0, R) \setminus \{z_0\}$ definen la misma función h , tiene que ser $a_{n_0} = \lim_{z \rightarrow z_0} h(z) = b_{n_0}$, y así tenemos una contradicción. ■

Definición 7.22 La serie de potencias dada por el teorema anterior se llama *serie de Taylor* de la función holomorfa f alrededor del punto z_0 .

Volviendo al ejemplo precedente (con $z_0 = 0$), la función $1/(1 + z^2)$ es holomorfa en $D(0, 1)$, por lo que su serie de Taylor en 0 tiene que converger en dicho disco y, como no puede ser holomorfa en ningún disco mayor (porque tiende a ∞ en $\pm i$), la serie no puede converger en ningún disco mayor, su radio de convergencia es necesariamente $R = 1$.

En cambio, la serie de Taylor en $z_0 = 0$ de una función entera, como e^z , $\sin z$, etc., tiene que converger necesariamente en todo el plano complejo, pues el teorema anterior es aplicable (con la misma serie) a cualquier disco $D(0, R)$, luego tiene que converger en todos ellos, es decir, en todo \mathbb{C} .

A su vez, los desarrollos en serie de Taylor se traducen en más propiedades de las funciones holomorfas. Veamos una muy simple:

Teorema 7.23 (Teorema de Liouville) *Toda función entera acotada es constante.*

DEMOSTRACIÓN: Si $f : \mathbb{C} \rightarrow \mathbb{C}$ es una función entera, por el teorema anterior admite un desarrollo en serie de Taylor

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$$

convergente en todo el plano complejo.

Supongamos que $f(z) \leq M$ para todo $z \in \mathbb{C}$. Por las fórmulas de Cauchy, para todo $r > 0$, se cumple que

$$\left| \frac{f^{(n)}(0)}{n!} \right| = \left| \frac{1}{2\pi i} \int_{|\zeta|=r} \frac{f(\zeta)}{\zeta^{n+1}} d\zeta \right| \leq \frac{M}{r^n},$$

donde hemos usado la propiedad c) del teorema 7.6. Si $n \geq 1$, haciendo tender r a infinito queda que $f^{(n)}(0) = 0$, luego la serie se reduce a $f(z) = f(0)$. ■

Veamos una aplicación:

Teorema 7.24 (Teorema fundamental del álgebra) *Todo polinomio no constante en $\mathbb{C}[z]$ tiene al menos una raíz compleja.*

DEMOSTRACIÓN: Sea $P(z)$ un polinomio no nulo. Si $P(z)$ no tuviera raíces complejas, la función $f(z) = 1/P(z)$ sería entera y $\lim_{z \rightarrow \infty} f(z) = 0$, luego f estaría acotada y, por el teorema anterior, sería constante, luego $P(z)$ también. ■

Una de las consecuencias más notables de los desarrollos en serie de Taylor se basa en esta propiedad de las series de potencias:

Teorema 7.25 *Sea $f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n$ una serie de potencias convergente en un disco $D(z_0, r)$ y supongamos que existe una sucesión $\{c_n\}_{n=0}^{\infty}$ contenida en $D(z_0, r) \setminus \{z_0\}$, convergente a z_0 y tal que $f(c_n) = 0$. Entonces f es idénticamente nula.*

DEMOSTRACIÓN: Como f es continua en z_0 , tenemos que

$$a_0 = f(z_0) = \lim_n f(c_n) = 0.$$

Vamos a probar que todos los a_n son nulos. En caso contrario, sea $n_0 > 0$ el menor índice tal que $a_{n_0} \neq 0$. Entonces

$$f(z) = (z - z_0)^{n_0} \sum_{n=n_0}^{\infty} a_n(z - z_0)^{n-n_0}.$$

La serie

$$g(z) = \sum_{n=n_0}^{\infty} a_n(z - z_0)^{n-n_0}$$

converge en $D(z_0, r)$ a una función holomorfa y $g(z_0) = a_{n_0} \neq 0$. Por la continuidad, existe $0 < r_0 < r$ tal que $g(z) \neq 0$ en $D(z_0, r_0)$, pero entonces $f(z)$ no se anula en ningún punto de $D(z_0, r_0) \setminus \{z_0\}$, cuando este disco debe contener infinitos términos de la sucesión c_n , donde f se anula, con lo que tenemos una contradicción. ■

Ahora necesitamos una definición:

Definición 7.26 Un abierto $A \subset \mathbb{C}$ es *conexo* si cuando $a, b \in A$ existen puntos z_0, \dots, z_n tales que $a = z_0$, $b = z_n$ y cada segmento $[z_{i-1}, z_i]^*$ está contenido en A .

Por ejemplo, es claro que todo abierto estrellado es conexo. Si un abierto no es conexo, es que está formado por dos o más “piezas” incomunicadas (por ejemplo, dos círculos disjuntos), de modo que una función definida sobre él consta en realidad de varias funciones independientes sin ninguna relación entre sí. Por eso en el teorema siguiente es necesario suponer que el dominio de las funciones es conexo:

Teorema 7.27 (Principio de prolongación analítica) Sean dos funciones holomorfas $f, g : A \subset \mathbb{C} \rightarrow \mathbb{C}$ definidas en un abierto conexo tales que existe una sucesión $\{c_n\}_{n=0}^{\infty}$ contenida en A , convergente a un punto $z_0 \in A$ de modo que $f(c_n) = g(c_n)$ para todo n . Entonces $f = g$.

DEMOSTRACIÓN: Sea $h = f - g$. Entonces h es una función holomorfa en A con la propiedad de que $h(c_n) = 0$ para todo n , y basta probar que h es idénticamente nula.

Como A es abierto, existe $r > 0$ tal que $D(z_0, r) \subset A$. La serie de Taylor de h alrededor de z_0 converge a h en este disco, y por el teorema 7.25 resulta que h es idénticamente nula en él.

Veamos que h es idénticamente nula en todo A , para lo cual⁷ tomamos un punto arbitrario $c \in A$. Por la conexión existen puntos $z_0, \dots, z_n \in A$ tales que $z_n = c$ y cada segmento $[z_{i-1}, z_i]^*$ está contenido en A . Eliminando repeticiones, podemos suponer que cada z_{i-1} es distinto de z_i e, intercalando un punto entre z_0 y z_1 , podemos suponer que $[z_0, z_1]^* \subset D(z_0, r)$, con lo que h es nula en todos los puntos de $[z_0, z_1]^*$. Basta probar que si h es nula en $[z_{i-1}, z_i]^*$, también lo es en $[z_i, z_{i+1}]^*$, pues así llegamos a que es nula en el último segmento, luego en particular en c .

En efecto, sea $r > 0$ tal que $D(z_i, r) \subset A$. Entonces h admite un desarrollo en serie de Taylor alrededor de z_i convergente en $D(z_i, r)$ y claramente $[z_{i-1}, z_i]^*$ contiene una sucesión de puntos convergente a z_i donde h se anula, luego por el teorema 7.25 tenemos que h es idénticamente nula en $D(z_i, r)$.

Llamemos $\phi = [z_i, z_{i+1}]$, es decir, $\phi(t) = (1-t)z_i + tz_{i+1}$. Basta probar que $h(\phi(t)) = 0$ para todo $t \in [0, 1]$. En caso contrario sea

$$\tau = \inf\{t \in [0, 1] \mid \phi(t) \neq 0\}.$$

Tienes que ser $\tau > 0$, pues por continuidad existe un $\delta > 0$ tal que si $|t| < \delta$, entonces $|\phi(t) - z_i| < r$, luego $\phi(t) \in D(z_i, r)$ y $h(\phi(t)) = 0$. Esto significa que δ es una cota inferior del conjunto cuyo ínfimo es τ , luego $\tau \geq \delta > 0$.

Sea $r' > 0$ tal que $D(\phi(\tau), r') \subset A$. De nuevo h admite un desarrollo en serie de Taylor en este disco y una sucesión en $[0, \tau[$ convergente a τ da lugar a una sucesión de puntos en dicho disco convergente a $\phi(\tau)$ en la que h se anula, luego de nuevo por el teorema 7.25 resulta que h se anula en todo el disco.

⁷El lector familiarizado con la topología podrá simplificar drásticamente el argumento que sigue demostrando (con la misma técnica) que el conjunto de puntos donde h se anula es abierto, y también cerrado por continuidad, por lo que tiene que ser todo A .

En particular $h(\phi(\tau)) = 0$, luego $\tau < 1$ (o de lo contrario sería $h(\phi(t)) = 0$ para todo t). Por continuidad existe un $\delta > 0$ tal que si $\tau < t < \tau + \delta$, entonces se cumple $\phi(t) \in D(\phi(\tau), r')$, luego $h(\phi(t)) = 0$, pero esto significa que $\tau + \delta$ es una cota inferior del conjunto cuyo ínfimo es τ , luego tendría que ser $\tau + \delta \leq \tau$, lo cual es absurdo. ■

El teorema anterior es uno de los que mejor ilustran la similitud entre las funciones holomorfas y los polinomios. Como un polinomio no nulo sólo puede tener un número finito de raíces, tenemos que si dos funciones polinómicas coinciden en infinitos puntos, necesariamente son iguales. El teorema anterior afirma esto mismo, pero con una hipótesis ligeramente más fuerte: dos funciones holomorfas (sobre un abierto conexo) que coinciden en una sucesión convergente son iguales.

El requisito de la convergencia es necesario, pues, por ejemplo, la función entera $\sin 2\pi z$ se anula en los números enteros, pero no coincide con la función nula, que también lo hace.

Así pues, si dos funciones holomorfas coinciden en cualquier pequeña región del plano complejo que contenga una sucesión convergente, por ejemplo, un segmento diminuto, o un arco de curva, necesariamente son iguales. En teoría, cualquier porción mínima de una función holomorfa la determina unívocamente sobre todo su dominio.

En particular ahora podemos afirmar que las funciones complejas e^z , $\sin z$, $\cos z$ son las únicas extensiones holomorfas de las funciones reales correspondientes. Veamos otra aplicación:

La función dseta de Riemann Consideremos las funciones

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}, \quad \eta(z) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^z}.$$

Ambas son absolutamente convergentes en el semiplano $\operatorname{Re} z > 1$, pero la segunda [ITAn 8.31] converge condicionalmente en el semiplano $\operatorname{Re} z > 0$. Además, para $\operatorname{Re} z > 1$ satisfacen la relación [ITAn (8.8)]

$$\eta(z) = (1 - 2^{1-z})\zeta(z).$$

El primer factor se anula cuando $2^{1-z} = 1$, que equivale a

$$2^{z-1} = e^{(z-1)\log 2} = 1.$$

Si $z = x + yi$, la ecuación $e^z = 1$ equivale a $e^x(\cos y + i \sin y) = 1$, que se cumple cuando $x = 0$ e $y = 2k\pi$, para cada entero k , es decir, cuando $z = 2k\pi i$. Por lo tanto, $e^{(z-1)\log 2} = 1$ se cumple cuando $(z-1)\log 2 = 2k\pi i$, es decir, cuando z toma uno de los valores

$$z_k = 1 + \frac{2k\pi i}{\log 2}.$$

Por lo tanto, la función

$$f(z) = \frac{\eta(z)}{1 - 2^{1-z}}$$

Es holomorfa en todo el semiplano $\operatorname{Re} z > 0$ salvo a lo sumo en los puntos z_k , donde se anula el denominador, y coincide con $\zeta(z)$ en el semiplano $\operatorname{Re} z > 0$.

Vemos así que la función dseta de Riemann admite una prolongación analítica al semiplano $\operatorname{Re} z > 0$ que no está definida en los puntos z_k , todos ellos sobre la recta $\operatorname{Re} z = 1$. El principio de prolongación analítica garantiza que esa prolongación no es “una prolongación”, sino “la prolongación”, por lo que podemos llamar $\zeta(z)$ a la extensión que hemos obtenido.

Ahora bien, por otra parte podemos considerar la serie

$$\eta^*(z) = \frac{1}{1^z} + \frac{1}{2^z} - \frac{2}{3^z} + \frac{1}{4^z} + \frac{1}{5^z} - \frac{2}{6^z} + \dots$$

que también converge en el semiplano $\operatorname{Re} z > 0$ por el teorema [ITAn 8.31], mientras que en el semiplano $\operatorname{Re} z > 1$ converge absolutamente (por ejemplo, porque está acotada por $2\zeta(z)$). Y ahora se cumple la relación

$$\eta^*(z) = (1 - 3^{1-z})\zeta(z).$$

En efecto, la convergencia absoluta permite reordenar las series, luego

$$\begin{aligned} \eta^*(z) &= -2 \sum_{n=1}^{\infty} \frac{1}{(3n)^z} + \sum_{n=1}^{\infty} \frac{1}{(3n+1)^z} + \sum_{n=1}^{\infty} \frac{1}{(3n+2)^z}, \\ \zeta(z) &= \sum_{n=1}^{\infty} \frac{1}{(3n)^z} + \sum_{n=1}^{\infty} \frac{1}{(3n+1)^z} + \sum_{n=1}^{\infty} \frac{1}{(3n+2)^z}, \end{aligned}$$

luego

$$\zeta(z) - \eta^*(z) = 3 \frac{1}{3^z} \sum_{n=1}^{\infty} \frac{1}{n^z} = 3^{1-z} \zeta(z).$$

Por lo tanto

$$g(z) = \frac{\eta^*(z)}{1 - 3^{1-z}}$$

es también una función holomorfa en el semiplano $\operatorname{Re} z > 0$ salvo a lo sumo en los puntos

$$z_k^* = 1 + \frac{2k\pi i}{\log 3}$$

que extiende a la función $\zeta(z)$. Y en este punto es crucial que, por el principio de prolongación analítica, ambas extensiones son la misma. Pero $z_k = z_l^*$ equivale a que

$$\frac{k}{\log 2} = \frac{l}{\log 3},$$

o también a que $\log 3^k = \log 2^l$, o simplemente a que $3^k = 2^l$, lo cual sólo es posible si $k = l = 0$. Por lo tanto, si $k \neq 0$, tenemos que la extensión está definida en z_k y en z_k^* , sea por una de las extensiones o por la otra, y así podemos concluir:

Teorema 7.28 *La función dseta de Riemann, definida en principio en el semiplano $\operatorname{Re} z > 1$ mediante*

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z},$$

admite una extensión holomorfa al semiplano $\operatorname{Re} z > 0$ salvo el punto $z_0 = 1$.

En [ITAn] vimos que

$$\lim_{x \rightarrow 1^+} \zeta(x) = +\infty,$$

por lo que la extensión de $\zeta(z)$ no puede ser holomorfa en $z_0 = 1$ (o el límite anterior debería valer $\zeta(1)$).

En realidad, aunque no lo vamos a probar aquí, es posible extender la función dseta a una función holomorfa en $\mathbb{C} \setminus \{1\}$, y esta extensión representa un papel central en el estudio de la distribución de los números primos, aunque ahora no estamos en condiciones de dar detalles al respecto.

Logaritmos complejos Para terminar probamos que toda función holomorfa que no se anule en un abierto estrellado admite un logaritmo holomorfo:

Teorema 7.29 *Sea $f : A \subset \mathbb{C} \rightarrow \mathbb{C}$ una función holomorfa en un abierto estrellado A que no se anule en ningún punto. Entonces existe una función holomorfa $g : A \rightarrow \mathbb{C}$ tal que $f(z) = e^{g(z)}$ para todo $z \in A$.*

DEMOSTRACIÓN: Como f no se anula, la función f/f' es holomorfa en A , luego el teorema 7.14 nos da una función holomorfa $g : A \rightarrow \mathbb{C}$ tal que $g' = f/f'$. Sea $h(z) = e^{g(z)}/f(z)$. Entonces

$$h'(z) = \frac{e^{g(z)} f'(z) - e^{g(z)} f'(z)}{f(z)^2} = 0,$$

Pero una función holomorfa con derivada nula en un abierto estrellado es constante (por ejemplo, por la regla de Barrow 7.8). Así pues, $f(z) = k e^{g(z)}$, para cierta constante $k \neq 0$. Podemos expresar $k = e^{k_0}$, y entonces, cambiando $g(z)$ por $k_0 + g(z)$, tenemos que $f(z) = e^{g(z)}$. ■

7.4 El teorema de Cauchy

En esta sección demostraremos el teorema 7.13, cuya prueba hemos dejado pendiente en la sección anterior y que, como hemos visto, está en la base de todos los resultados que hemos obtenido sobre las funciones holomorfas. En primer lugar necesitamos algunos hechos elementales sobre triángulos.

Definimos el diámetro de un triángulo $T = (a, b, c)$ como el diámetro de $[T]$, es decir, como

$$\delta(T) = \sup\{|z - z'| \mid z, z' \in [T]\}.$$

Se cumple que $\delta(T) \leq L(T)$. En efecto, tenemos que $|c - a| \leq |c - b| + |b - a|$, luego

$$2|c - a| \leq |c - b| + |b - a| + |c - a| = L(T).$$

Esto prueba que $c \in D = \overline{D(a, L(T)/2)}$, y lo mismo le ocurre a b (y trivialmente a a), pero D es convexo, luego $[T] \subset D$, luego $\delta(T) \leq \delta(D) = L(T)$. ■

Por último, observemos que $[T]$ es compacto. Obviamente es acotado, pero falta probar que es cerrado. Para ello tomamos una sucesión de puntos

$$z_n = \alpha_n a + \beta_n b + \gamma_n c \in [T]$$

que converja a un punto z , y tenemos que probar que $z \in [T]$. Como $\alpha_n \in [0, 1]$, que es un conjunto compacto, tomando una subsucesión, podemos suponer que $\{\alpha_n\}_{n=0}^\infty$ converge a un $\alpha \in [0, 1]$, e igualmente podemos suponer que $\{\beta_n\}_{n=0}^\infty$ converge a β y $\{\gamma_n\}_{n=0}^\infty$ converge a γ . Como $\alpha_n + \beta_n + \gamma_n = 1$, al tomar límites queda que $\alpha + \beta + \gamma = 1$, y tomando límites en la definición de z_n queda que $z = \alpha a + \beta b + \gamma c$, luego $z \in [T]$. ■

Con esto ya podemos demostrar el teorema de Cauchy. Pongamos que el triángulo dado tiene vértices $T = (a, b, c)$. Distinguiamos tres casos:

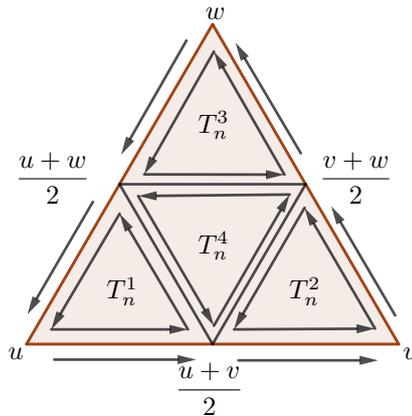
1) $p \notin [T]$.

Vamos a demostrar que existe una sucesión $\{T_n\}_{n=0}^\infty$ de triángulos tal que $T_0 = T$ y además

a) $[T_{n+1}] \subset [T_n] \subset A$.

b) $L(T_{n+1}) = L(T_n)/2$.

c) $\frac{1}{4} \left| \int_{T_n} f(\zeta) d\zeta \right| \leq \left| \int_{T_{n+1}} f(\zeta) d\zeta \right|$.



Supongamos construidos T_0, \dots, T_n y sea $T_n = [u, v, w]$. Definimos los triángulos

$$T_n^1 = \left[u, \frac{u+v}{2}, \frac{u+w}{2} \right], \quad T_n^2 = \left[\frac{u+v}{2}, v, \frac{v+w}{2} \right],$$

$$T_n^3 = \left[\frac{v+w}{2}, w, \frac{u+w}{2} \right], \quad T_n^4 = \left[\frac{u+w}{2}, \frac{u+v}{2}, \frac{v+w}{2} \right].$$

Claramente los vértices de los triángulos T_n^i están contenidos en $[T_n]$. Como éste es convexo, $[T_n^i] \subset [T_n]$. También es obvio que $L(T_n^i) = L(T_n)/2$. Esto significa que cualquier T_n^i sirve como T_{n+1} en lo que respecta a las condiciones a) y b). Basta probar que al menos uno de ellos cumple c). Consideremos la suma

$$\sum_{i=1}^4 \int_{T_n^i} f(\zeta) d\zeta.$$

Al descomponer cada integral como suma de las integrales sobre cada uno de los lados del triángulo vemos que la función se integra dos veces sobre los lados de T_n^4 , pero en sentidos opuestos (véase la figura), luego estas integrales se cancelan, y lo que queda es la integral sobre los lados de T_n . Así pues,

$$\sum_{i=1}^4 \int_{T_n^i} f(\zeta) d\zeta = \int_{T_n} f(\zeta) d\zeta.$$

Por consiguiente

$$\left| \int_{T_n} f(\zeta) d\zeta \right| \leq \sum_{i=1}^4 \left| \int_{T_n^i} f(\zeta) d\zeta \right|,$$

lo que implica que al menos una de las integrales de la derecha ha de ser mayor o igual que la cuarta parte del miembro izquierdo, luego, tomando como T_{n+1} el triángulo correspondiente a dicha integral, se cumple c).

Con esto tenemos probada la existencia de la sucesión $\{T_n\}_{n=0}^{\infty}$. Vamos a probar que existe un punto z_0 que está en todos los triángulos $[T_n]$.

Tomamos $w_n \in [T_n]$, para cada n . Como T_0 es compacto, la sucesión tiene una subsucesión $\{w_{n_k}\}_{k=0}^{\infty}$ convergente a un punto z_0 . Para cada n , existe un k_0 tal que $n_{k_0} \geq n$ y si $k \geq k_0$ entonces $w_{n_k} \in [T_{n_k}] \subset [T_{n_{k_0}}] \subset [T_n]$. Como $[T_n]$ es cerrado, esto implica que $z_0 \in [T_n]$.

Como $z_0 \in [T]$ y estamos suponiendo que $p \notin [T]$, se cumple que $z_0 \neq p$, luego f es derivable en z_0 . En consecuencia, dado $\epsilon > 0$, existe un $\delta > 0$ tal que si $|z - z_0| < \delta$, entonces

$$\left| \frac{f(z) - f(z_0)}{z - z_0} - f'(z_0) \right| < \epsilon,$$

luego

$$|f(z) - f(z_0) - f'(z_0)(z - z_0)| < \epsilon|z - z_0|.$$

Por la propiedad b) tenemos que $\delta([T_n]) \leq L(T_n) = 2^{-n}L(T)$. Como los diámetros tienden a 0, existe un n tal que $[T_n] \subset D(z_0, \delta)$. La función

$$-f(z_0) - f'(z_0)(z - z_0)$$

es un polinomio, luego tiene primitiva, luego, por el teorema 7.8,

$$\int_{T_n} (-f(z_0) - f'(z_0)(\zeta - z_0)) d\zeta = 0.$$

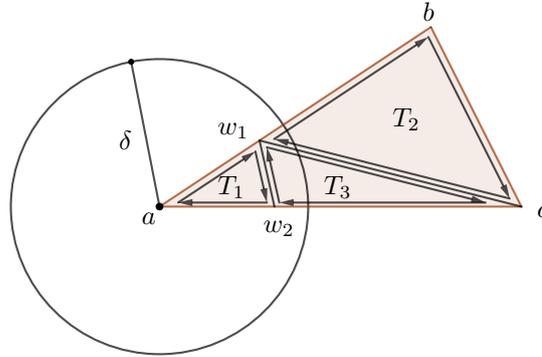
Teniendo en cuenta esto y la propiedad c), resulta

$$\begin{aligned} \left| \int_T f(\zeta) d\zeta \right| &\leq 4^n \left| \int_{T_n} f(\zeta) d\zeta \right| = 4^n \left| \int_{T_n} (f(\zeta) - f(z_0) - f'(z_0)(\zeta - z_0)) d\zeta \right| \\ &< 4^n \epsilon \int_{T_n} |\zeta - z_0| d\zeta \leq 4^n \epsilon L(T_n) \delta([T_n]) \leq 4^n \epsilon L(T_n)^2 \leq \epsilon L(T)^2, \end{aligned}$$

donde hemos usado la propiedad c) del teorema 7.6. Como esto es cierto para todo $\epsilon > 0$ concluimos que la integral es nula.

2) Supongamos ahora que p es uno de los vértices a, b, c de T . No perdemos generalidad si suponemos que $p = a$.

Sea $\delta_0 > 0$ tal que $\overline{D(a, \delta_0)} \subset A$. Como el disco es compacto y f es continua en él, existe un $M > 0$ tal que $|f(z)| \leq M$ siempre que $|z - a| \leq \delta_0$. Fijemos ahora $0 < \delta < \delta_0$ y tomemos puntos w_1 y w_2 en los segmentos $[a, b]^*$ y $[a, c]^*$ tales que $|w_1 - a| < \delta$, $|w_2 - a| < \delta$. Consideramos los triángulos T_1 , T_2 y T_3 que muestra la figura:



Como antes, al descomponer las integrales sobre T_1 , T_2 y T_3 en las sumas de las integrales sobre sus lados, las integrales correspondientes a los lados interiores se cancelan y queda que

$$\int_T f(\zeta) d\zeta = \sum_{i=1}^3 \int_{T_i} f(\zeta) d\zeta.$$

Por el caso 1) las integrales sobre T_2 y T_3 son nulas, luego

$$\int_T f(\zeta) d\zeta = \int_{T_1} f(\zeta) d\zeta,$$

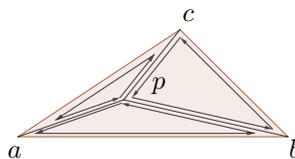
luego

$$\left| \int_T f(\zeta) d\zeta \right| = \left| \int_{T_1} f(z) dz \right| \leq ML(T_1) \leq M(|w_1 - a| + |w_2 - a| + |w_1 - w_2|) \leq 4M\delta$$

para todo $\delta > 0$, luego la integral es nula.

3) Por último, consideramos el caso en que $p \in [T]$, pero no es un vértice.

Dividimos T en tres triángulos como indica la figura. Una vez más la integral de f sobre T se descompone en suma de las integrales sobre los tres triángulos, y las tres integrales son nulas por el caso 2). ■

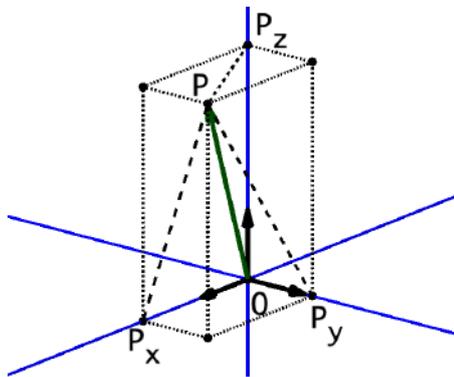


Apéndice A

Geometría analítica tridimensional

Recordemos que una recta graduada es una recta en la que hemos seleccionado un origen P_0 y un punto unitario P_1 , lo cual da lugar a una correspondencia biunívoca entre sus puntos y los números reales. La longitud $\overline{P_0P_1}$ es la *unidad de longitud* establecida por la graduación.

Puntos y vectores Si elegimos un punto O del espacio como *origen de coordenadas* y trazamos en él tres *ejes* (rectas) perpendiculares dos a dos y consideramos en ellos graduaciones con la misma unidad de longitud, tenemos un *sistema de referencia*, el cual permite establecer una correspondencia biunívoca entre los puntos P del espacio y las ternas (x, y, z) de números reales:



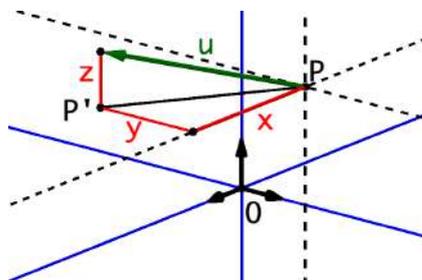
Para ello llamamos P_x, P_y, P_z a los pies de las perpendiculares por P en los tres ejes, de modo que (x, y, z) es la terna de números reales correspondientes a dichos pies en las graduaciones respectivas de los ejes, es decir, $x = \pm \overline{OP_x}$, donde el signo depende de si P_x está en la semirrecta respecto de O del punto unitario o en la semirrecta opuesta, e igualmente con las otras coordenadas.

Recíprocamente, cualquier terna (x, y, z) determina puntos P_x, P_y, P_z en los ejes, y el plano paralelo al plano (que contiene a los ejes) YZ y que pasa por P_x , el plano paralelo al plano (que contiene a los ejes) XZ que pasa por P_y y el plano paralelo al plano (que contiene a los ejes) XY que pasa por P_z se cortan en un punto P cuyas coordenadas son (x, y, z) .

Nota Existe un convenio universalmente aceptado sobre la *orientación* de los ejes en una representación gráfica conocido como “regla del pulgar” o, más precisamente, como “regla de la mano derecha”, según el cual, si apuntamos con el índice de la mano derecha hacia el semieje X positivo y con el dedo medio hacia el semieje Y positivo, el pulgar debe apuntar hacia el semieje Z positivo. La figura anterior muestra los puntos de coordenadas $(1, 0, 0)$, $(0, 1, 0)$ y $(0, 0, 1)$, que señalan los semiejes positivos y podemos comprobar que cumple este criterio. ■

Si llamamos \mathbb{R}^3 al conjunto de todas las ternas (x, y, z) de números reales, acabamos de mostrar que, fijando un sistema de referencia, podemos identificar los puntos del espacio con los elementos de \mathbb{R}^3 .

Sin embargo, a menudo convendrá considerar a las ternas de \mathbb{R}^3 , no como puntos, sino como “instrucciones” para movernos de un punto a otro. En tal caso nos referiremos a las ternas como “*vectores*”. Así, $u = (3, -2, 1)$ puede verse como las instrucciones que nos dicen que, a partir de un punto cualquiera P , nos movamos hasta el punto que resulta de sumar 3 unidades a su coordenada x (es decir, movernos tres unidades paralelamente al eje X en sentido positivo), restar 2 unidades al eje y (movernos 2 unidades paralelamente al eje Y en sentido negativo) y sumar 1 unidad a su coordenada z (movernos 1 unidad paralelamente al eje Z en sentido positivo).



El punto Q al que llegamos se representa por $Q = P + u$.

Más en general, en \mathbb{R}^3 podemos definir una suma y un producto por un número real dados por:

$$(x, y, z) + (x', y', z') = (x + x', y + y', z + z'), \quad \lambda(x, y, z) = (\lambda x, \lambda y, \lambda z)$$

y acabamos de dar una interpretación geométrica a la suma de un punto P más un vector u : es el punto $Q = P + u$ al que se llega cuando aplicamos a P las instrucciones de movimiento correspondientes a v .

A su vez, si P y Q son puntos cualesquiera, entonces $u = Q - P$ es el vector que, sumado a P , nos lleva a Q , lo cual nos da una interpretación geométrica a la resta de dos puntos.

Si llamamos P' al punto al que llegamos cuando sólo modificamos las coordenadas dos primeras coordenadas de P (véase la figura anterior), dos aplicaciones del teorema de Pitágoras¹ nos dan que

$$\overline{PQ}^2 = \overline{PP'}^2 + z^2 = x^2 + y^2 + z^2.$$

Por lo tanto, si definimos la *norma* de un vector $u = (x, y, z)$ de \mathbb{R}^3 como

$$\|u\| = \sqrt{x^2 + y^2 + z^2},$$

hemos probado que $\|u\|$ es la longitud de cualquier segmento de extremos P y $P+u$, donde P es un punto arbitrario. Podemos expresar esto diciendo que $\|u\|$ es la longitud del vector u . Equivalentemente, hemos probado que la longitud de un segmento PQ es $\|Q - P\|$.

Teorema A.1 *La recta que pasa por dos puntos distintos P y Q está formada por los puntos de la forma*

$$X = P + \lambda(Q - P), \quad \lambda \in \mathbb{R}.$$

Los puntos del segmento \overline{PQ} son los correspondientes a $0 \leq \lambda \leq 1$.

DEMOSTRACIÓN: Sea $X_\lambda = P + \lambda(Q - P)$. Es claro que si $\lambda = 0, 1$ el punto X_λ está en la recta (porque es P o Q). Supongamos que $0 < \lambda < 1$. Entonces

$$\|X_\lambda - P\| = \|\lambda(Q - P)\| = \lambda\|Q - P\|,$$

$$\|Q - X_\lambda\| = \|Q - P - \lambda(Q - P)\| = \|(1 - \lambda)(Q - P)\| = (1 - \lambda)\|Q - P\|,$$

luego

$$\|X_\lambda - P\| + \|Q - X_\lambda\| = \|Q - P\|,$$

y esto implica² que X_λ está entre P y Q , luego en particular en la recta que pasa por P y Q .

Si $\lambda > 1$ se prueba igualmente que Q está entre P y X_λ mientras que si $\lambda < 0$ concluimos que P está entre X_λ y Q .

Más aún, si λ es arbitrario, tenemos que

$$\|X_\lambda - P\| = |\lambda|\|Q - P\|,$$

luego, para cada número real $d > 0$, existen dos puntos de la forma X_λ que cumplen $\|X_\lambda - P\| = d$, a saber, los correspondientes a $\lambda = \pm d\|Q - P\|^{-1}$, y en una recta sólo hay dos puntos a una misma distancia de uno dado, luego concluimos que los puntos de la forma X_λ son todos los puntos de la recta que pasa por P y Q . ■

¹Hay que tratar aparte los casos en los que alguna de las coordenadas de u es nula, en los que se llega trivialmente a la misma conclusión.

²Si X_λ no estuviera en la recta que pasa por P y Q , la desigualdad triangular implicaría que la suma de las distancias a P y Q sería mayor que PQ y si está en la recta, pero no entre P y Q , se concluye también inmediatamente que no puede darse la igualdad.

Así pues, los puntos de una recta pueden expresarse en la forma

$$P + \lambda u,$$

donde P es uno de sus puntos y u es un vector no nulo, que recibe el nombre de *vector director* de la recta. No es único, pero dos vectores directores de una misma recta cumplirán necesariamente que $u_1 = \lambda u_2$, para cierto $\lambda \neq 0$.

En particular, si u es un vector no nulo, los puntos de la forma $\lambda u = O + \lambda u$ forman una recta que pasa por el origen de coordenadas O . Esto nos da una interpretación del producto λu :

Cuando se multiplica por λ un vector no nulo u , obtenemos otro vector con la misma dirección que u , pero cuya norma es $\|\lambda u\| = |\lambda| \|u\|$. Esto significa que multiplicar vectores por un mismo λ no nulo da lugar a vectores λ veces mayores (entendiendo que son menores si $|\lambda| < 1$) que tendrán el mismo sentido (es decir, que estarán en la misma semirrecta de origen O) si $\lambda > 0$ y sentido opuesto si $\lambda < 0$.

En otras palabras, si multiplicamos por λ los puntos de una figura obtenemos otra figura “a escala”, que será $|\lambda|$ veces mayor (entendiendo que será menor si $|\lambda| < 1$). Por ello, cuando se quiere hablar de números reales por oposición a “vectores”, se les llama “*escalares*”.

El producto escalar En \mathbb{R}^3 podemos definir un producto mediante

$$(x, y, z) \cdot (x', y', z') = xx' + yy' + zz'.$$

Se llama *producto escalar* porque el resultado no es un vector, sino un escalar. En estos términos tenemos que

$$\|u\| = \sqrt{u \cdot u}.$$

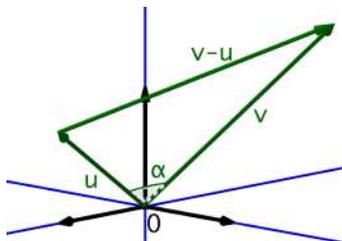
En general, el producto escalar tiene una interpretación geométrica muy simple. Si calculamos

$$(v - u) \cdot (v - u) = u \cdot u + v \cdot v - 2u \cdot v,$$

es decir,

$$\|v - u\|^2 = \|u\|^2 + \|v\|^2 - 2u \cdot v,$$

en el caso en que u y v son no nulos y no están sobre la misma recta, forman un triángulo con el origen O de coordenadas cuyos lados miden $\|u\|$, $\|v\|$ y $\|v - u\|$:



El teorema del coseno nos da que

$$\|v - u\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\|v\|\cos \alpha,$$

donde $\alpha = \widehat{u, v}$ es el ángulo que forman los vectores en O . Por lo tanto, concluimos que

$$u \cdot v = \|u\|\|v\|\cos \widehat{u, v}$$

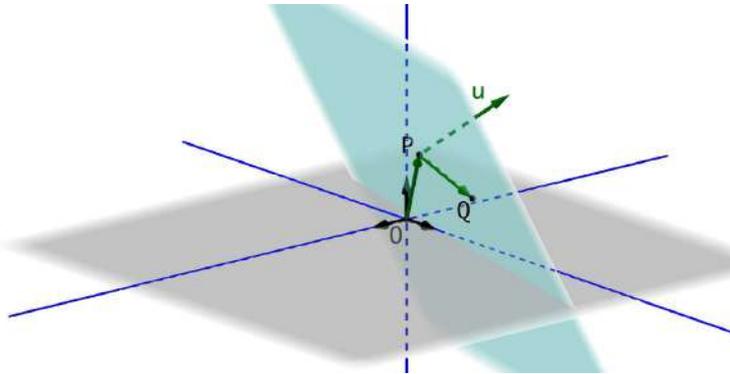
Si u y v son no nulos pero tienen la misma dirección, es decir, si $v = \lambda u$, para cierto escalar λ no nulo, la fórmula anterior sigue siendo válida si convenimos en que el ángulo que forman es $\widehat{u, v} = 0$ si $\lambda > 0$ o bien $\widehat{u, v} = \pi$ si $\lambda < 0$. En definitiva:

El producto escalar de dos vectores no nulos es igual al producto de sus normas por el coseno del ángulo que forman.

Definición A.2 Dos vectores u y v son *ortogonales* si $u \cdot v = 0$.

Si son no nulos, esto significa que $\cos \widehat{u, v} = 0$, luego u y v forman un ángulo recto. Por lo tanto, dos rectas secantes son perpendiculares si y sólo si sus vectores directores son ortogonales.

Planos Si un plano pasa por un punto P y la recta perpendicular al plano que pasa por P tiene vector director $u = (a, b, c)$, entonces un punto $Q = (x, y, z)$ está en el plano si y sólo si $Q - P$ es ortogonal a u :



Equivalentemente, Q está en el plano si y sólo si $u \cdot (Q - P) = 0$, que equivale a $u \cdot Q = u \cdot P$ o, llamando $d = u \cdot P$, la condición explícita es:

$$ax + by + cz = d.$$

Recíprocamente, los puntos que cumplen una ecuación de esta forma con $(a, b, c) \neq (0, 0, 0)$ forman el plano que pasa por cualquier punto P que cumpla la ecuación y que es perpendicular a la recta que pasa por P con vector director $u = (a, b, c)$. Pero esto prueba además que todas las rectas perpendiculares a un mismo plano admiten un mismo vector director $u = (a, b, c)$, independiente del

punto en el que lo cortan. Este vector está determinado salvo múltiplos por un escalar no nulo, y se llama *vector normal* al plano. El teorema siguiente resume lo que hemos obtenido:

Teorema A.3 *Todo plano Π está formado por los puntos que cumplen una ecuación de la forma*

$$ax + by + cz = d,$$

donde $u = (a, b, c)$ recibe el nombre de *vector normal* a Π , y es un *vector director común* a todas las rectas perpendiculares a Π .

Se dice que dos rectas son *paralelas* si coinciden o bien están contenidas en un mismo plano y no se cortan.

Teorema A.4 *Dos rectas son paralelas si y sólo si tienen el mismo vector director.*

DEMOSTRACIÓN: Consideremos dos rectas distintas con el mismo vector director u , que estarán formadas, respectivamente, por los puntos de la forma

$$P + \lambda u, \quad Q + \mu u,$$

donde necesariamente $P \neq Q$, o las rectas coincidirían. Sea $v = Q - P$ y sea Π el plano que contiene a la primera recta y a la recta PQ (dos rectas que se cortan en un punto P siempre están contenidas en un plano). Sea n un vector director de la recta perpendicular a Π por P . Entonces los puntos de Π son los puntos X que cumplen $n \cdot (X - P) = 0$, pero esto lo cumplen los puntos de la segunda recta, ya que

$$n \cdot (Q + \mu u - P) = n \cdot (Q - P) + \mu n \cdot (P + u - P) = 0,$$

ya que tanto Q como $P + u$ están en Π .

Así pues, las dos rectas están contenidas en el plano Π , y es claro que si se cortan coincidirían, pues serían dos rectas que pasan por el mismo punto con el mismo vector director. Por lo tanto son paralelas.

Recíprocamente, si dos rectas con vectores directores u y u' son paralelas, ambas son perpendiculares a un mismo plano, luego ambas tienen por vector director a cualquier vector normal a dicho plano. ■

Teorema A.5 *El plano que pasa por tres puntos no alineados P, Q, R está formado por los puntos de la forma*

$$P + \lambda(Q - P) + \mu(R - P), \quad \lambda, \mu \in \mathbb{R}.$$

DEMOSTRACIÓN: Sea n el vector normal al plano que pasa por P, Q, R . Entonces

$$n \cdot (Q - P) = n \cdot (R - P) = 0,$$

luego todo punto X de la forma indicada cumple

$$n \cdot (X - P) = \lambda n \cdot (Q - P) + \mu n \cdot (R - P) = 0$$

y, por consiguiente, está en el plano.

Recíprocamente, si X está en el plano, la paralela a PR por X cortará a PQ en un punto que será de la forma $P + \lambda(Q - P)$, y la paralela a PQ por X cortará a PR en un punto de la forma $P + \mu(R - P)$. Consideramos entonces el punto

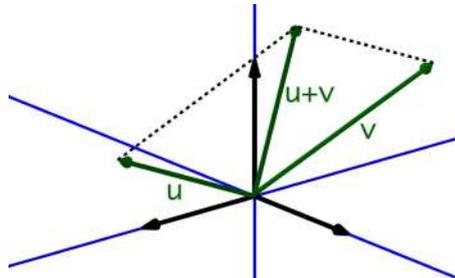
$$X' = P + \lambda(Q - P) + \mu(R - P).$$

La recta paralela a PR por X' está formada por los puntos de la forma

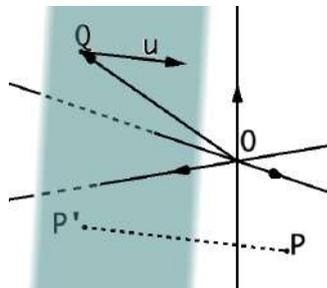
$$X' + \mu'(R - P) = P + \lambda(Q - P) + \mu(R - P) + \mu'(R - P), \quad \mu' \in \mathbb{R}$$

y corta a PQ cuando $\mu' = -\mu$ en el punto $P + \lambda(Q - P)$, luego coincide con la paralela a PR por X . Igualmente, la paralela a PQ por X' coincide con la paralela a PQ por X , luego $X = X'$, ya que ambos puntos son las intersecciones de dichas paralelas. ■

La suma de dos vectores u y v se puede interpretar como el vector $u + v$ que, al aplicárselo a un punto P nos lleva al punto que resulta de aplicar v al punto que resulta de aplicar u a P . La prueba del teorema anterior nos da una interpretación geométrica más directa: $u + v$ es el cuarto vértice del paralelogramo de vértices O , u y v :



Proyecciones ortogonales La proyección ortogonal de un punto P en un plano Π es el pie P' de la perpendicular a Π por P :



Si el plano Π pasa por un punto Q y tiene a u por vector normal, la recta perpendicular a Π por P tiene a u por vector director, por lo que existe un escalar λ tal que $P' = P + \lambda u$. Por otra parte, como P' está en Π , el vector $P' - Q$ tiene que ser ortogonal a u , luego $(P' - Q) \cdot u = 0$, es decir,

$$(P + \lambda u - Q) \cdot u = (P - Q) \cdot u + \lambda u \cdot u = 0,$$

luego

$$\lambda = -\frac{(P - Q) \cdot u}{u \cdot u},$$

y así la proyección ortogonal de P es

$$P' = P - \frac{(P - Q) \cdot u}{u \cdot u} u.$$

Un caso particular se da cuando el plano Π pasa por el origen O , en cuyo caso podemos tomar $Q = O$ y la fórmula se reduce a

$$P' = P - \frac{P \cdot u}{u \cdot u} u,$$

y en este caso vemos que la proyección ortogonal en Π es una aplicación lineal, es decir, que cumple

$$(\lambda P + \mu Q)' = \lambda P' + \mu Q'.$$

El producto vectorial Dados dos vectores u y v que no sean uno múltiplo del otro, no es inmediato cómo calcular un tercer vector w no nulo que sea ortogonal a ambos, es decir, tal que $u \cdot w = v \cdot w = 0$. Observemos que esta condición determina w salvo múltiplo por un escalar no nulo. Vamos a construir uno en concreto al que llamaremos $u \times v$ de modo que el producto así definido tenga propiedades algebraicas notables.

Definición A.6 Dados dos vectores u y v , definimos su *producto vectorial* $u \times v$ de modo que:

- a) Si uno de los vectores es múltiplo del otro, entonces $u \times v = 0$.
- b) En caso contrario, $u \times v$ es ortogonal a ambos,

$$\|u \times v\| = \|u\| \|v\| |\sin \widehat{u, v}|$$

y el sentido de $u \times v$ está determinado por la “regla de la mano derecha”, es decir, que si el índice de la mano derecha apunta en la dirección de u y el dedo medio apunta en la dirección de v , entonces $u \times v$ apunta en la dirección del pulgar.³

³Más precisamente, adoptamos esta definición si previamente hemos adoptado el convenio de orientar los ejes coordenados con el convenio indicado en la nota de la página 454, que en estos términos equivale a que $(0, 0, 1) = (1, 0, 0) \times (0, 1, 0)$. Si usáramos el convenio opuesto para los ejes, es decir, si $(0, 0, 1)$ tuviera el sentido opuesto al que exige la regla de la mano derecha, entonces tendríamos que definir el producto vectorial también con sentido opuesto, para que siguiera siendo cierto que $(0, 0, 1) = (1, 0, 0) \times (0, 1, 0)$. Vamos a ver que adoptando a la vez estos dos convenios el producto vectorial admite una definición puramente algebraica que es independiente de la regla de la mano derecha.

Observemos que $|\widehat{\text{sen } u, v}| = 0$ sucede exactamente cuando $\widehat{u, v} = 0, \pi$, lo que a su vez equivale a que u y v sean uno múltiplo del otro, por lo que la regla

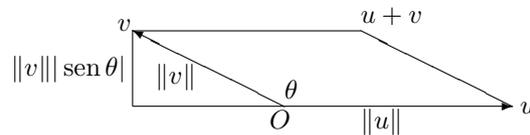
$$\|u \times v\| = \|u\| \|v\| |\widehat{\text{sen } u, v}|$$

se cumple en todos los casos, tanto si uno de los vectores es múltiplo del otro como si no, y $u \times v = 0$ si y sólo si uno de los vectores es múltiplo del otro (en particular si uno de ellos es nulo o si $u = v$). La regla de la mano derecha también implica que

$$u \times v = -v \times u,$$

por lo que el producto vectorial no es conmutativo.

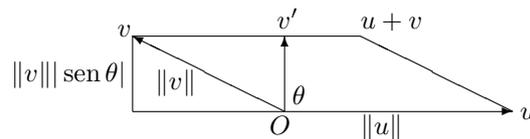
Una interpretación geométrica de la norma del producto vectorial es que $\|u \times v\|$ es el área del paralelogramo de vértices $0, u, v, u + v$:



Equivalentemente, el área del triángulo de vértices O, u, v es $(1/2)u \times v$.

Ahora observamos que, si $u \neq 0$, entonces $u \times v$ sólo depende de la proyección ortogonal v' de v en el plano que pasa por el origen con vector normal u , es decir, que $u \times v = u \times v'$.

En efecto, en la figura siguiente, en la que la página del libro se identifica con el plano que contiene a los vectores u y v , el vector v' es el que se indica, de modo que $\widehat{u, v'} = \pi/2$, luego $|\widehat{\text{sen } u, v'}| = 1$ y $\|v'\| = \|v\| \text{sen } \theta$. Esto implica que $\|u \times v\| = \|u \times v'\|$.



Si v es múltiplo de u , entonces $v' = 0$ y ambos productos vectoriales son nulos. En caso contrario, ambos son perpendiculares al plano de la página y la regla de la mano derecha nos da que ambos tienen el mismo sentido. Concretamente, en el caso de la figura ambos apuntan hacia el lector, y si el vector v estuviera en el semiplano inferior, entonces ambos apuntarían en la dirección opuesta al lector, luego en cualquier caso los dos vectores son iguales.

Este hecho nos permite probar una propiedad algebraica fundamental:

$$u \times (v_1 + v_2) = u \times v_1 + u \times v_2.$$

En efecto, la igualdad es trivial si u es nulo, así que podemos suponer que no lo es. En el apartado precedente hemos visto que la proyección ortogonal

sobre un plano que pase por O es lineal, de modo que $(v_1 + v_2)' = v_1' + v_2'$, y la observación precedente implica entonces que la fórmula que queremos probar equivale a

$$u \times (v_1' + v_2') = u \times v_1' + u \times v_2'.$$

Por lo tanto, no perdemos generalidad si suponemos que v_1 y v_2 son ortogonales a u , y lo mismo vale entonces para $v_1 + v_2$. A continuación observamos que si v es cualquier vector ortogonal a u , entonces $|\widehat{u, v}| = 1$, luego $\|u \times v\| = \|u\|\|v\|$. Más aún, si miramos el plano perpendicular a u de modo que u apunte hacia nosotros, la regla de la mano derecha hace que $u \times v$ sea el vector que resulta de girar v un ángulo $\pi/2$ en sentido antihorario y multiplicarlo por $\|u\|$, pero tanto el giro como la multiplicación por $\|u\|$ conservan la suma, luego se cumple la relación requerida.

El mismo argumento prueba que $u \times (\alpha v) = \alpha(u \times v)$, donde α es un escalar.

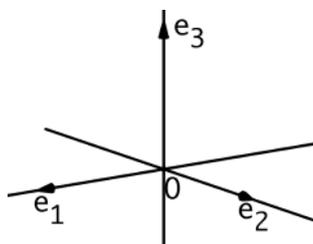
Finalmente podemos obtener una expresión puramente algebraica para el producto vectorial de dos vectores arbitrarios. Para ello observemos que, si

$$e_1 = (1, 0, 0), \quad e_2 = (0, 1, 0), \quad e_3 = (0, 0, 1),$$

con el convenio de la regla de la mano derecha adoptado tanto en la definición del producto vectorial como en la orientación de los ejes, se cumple que

$$e_1 \times e_2 = e_3, \quad e_2 \times e_3 = e_1, \quad e_3 \times e_1 = e_2,$$

como el lector puede comprobar con su mano derecha a partir de la figura siguiente:



Por consiguiente, si

$$u = (u_1, u_2, u_3) = u_1 e_1 + u_2 e_2 + u_3 e_3, \quad v = (v_1, v_2, v_3) = v_1 e_1 + v_2 e_2 + v_3 e_3,$$

entonces

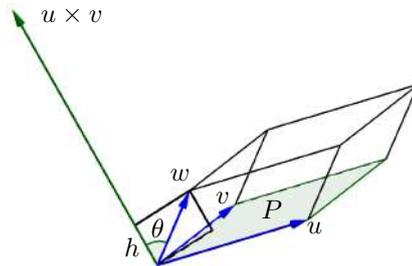
$$\begin{aligned} u \times v &= u_1 v_1 e_1 \times e_1 + u_1 v_2 e_1 \times e_2 + u_1 v_3 e_1 \times e_3 \\ &= + u_2 v_1 e_2 \times e_1 + u_2 v_2 e_2 \times e_2 + u_2 v_3 e_2 \times e_3 \\ &= + u_3 v_1 e_3 \times e_1 + u_3 v_2 e_3 \times e_2 + u_3 v_3 e_3 \times e_3 \\ &= u_1 v_2 e_3 - u_1 v_3 e_2 - u_2 v_1 e_3 + u_2 v_3 e_1 + u_3 v_1 e_2 - u_3 v_2 e_1 \\ &= (u_2 v_3 - u_3 v_2) e_1 + (u_3 v_1 - u_1 v_3) e_2 + (u_1 v_2 - u_2 v_1) e_3. \end{aligned}$$

Así pues, podemos tomar como definición de producto vectorial de vectores la fórmula puramente algebraica

$$u \times v = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} e_1 - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} e_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} e_3,$$

que no depende para nada de la regla de la mano derecha, pero el desarrollo precedente nos da su interpretación geométrica.

El producto mixto Consideremos ahora tres vectores cualesquiera, u, v, w . Hemos visto que el producto $u \times v$ es un vector ortogonal a u y v cuyo módulo es el área del paralelogramo P determinado por u y v y cuyo sentido viene determinado por la “regla de la mano derecha”.



A su vez, $(u \times v) \cdot w = \|u \times v\| \|w\| \cos \theta$, donde θ es el ángulo que forman $u \times v$ y w . El coseno será positivo o negativo según si θ es menor o mayor que $\pi/2$, y es 0 cuando w es perpendicular a $u \times v$, es decir, cuando w está en el plano determinado por u y v . Ahora observamos que $h = \|w\| \cos \theta$ es la altura del paralelepípedo determinado por u, v, w , por lo que, de acuerdo con lo visto en la sección 2.6, resulta que $|(u \times v) \cdot w| = \|u \times v\| h$ es el volumen de dicho paralelepípedo.

El signo de $(u \times v) \cdot w$ será positivo si w forma un ángulo menor que $\pi/2$ con $u \times v$, lo cual equivale a que w apunte hacia el mismo semiplano que $u \times v$ respecto del plano determinado por u y v , es decir, si w apunta hacia el semiplano determinado por u y v según la “regla de la mano derecha”.

Definición A.7 Definimos el *producto mixto* de tres vectores u, v, w como

$$(u, v, w) = (u \times v) \cdot w = u \cdot (v \times w).$$

La segunda igualdad se debe a que, en valor absoluto, ambos términos representan el volumen del paralelepípedo determinado por u, v y w , y el lector puede comprobar sin más que mirar la figura precedente que w apunta hacia el semiplano determinado por u y v mediante la regla de la mano derecha si y sólo si u apunta hacia el semiplano determinado por v y w .

Equivalentemente, esta la igualdad puede expresarse como:

$$(u, v, w) = (v, w, u),$$

de modo que *el producto mixto no se altera si se permutan cíclicamente los tres vectores*.

Por otra parte, teniendo en cuenta que $u \times v = -v \times u$, vemos que

$$(u, v, w) = -(v, u, w), \quad (u, v, w) = -(u, w, v)$$

y también

$$(u, v, w) = (w, u, v) = -(u, w, v) = -(w, v, u),$$

luego el producto mixto cambia de signo siempre que se intercambian dos de los vectores. ■

Determinantes 3×3 Para calcular explícitamente productos mixtos es preferible definir el *determinante* de una matriz 3×3

$$A = \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$$

como

$$\det A = |A| = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = (u, v, w),$$

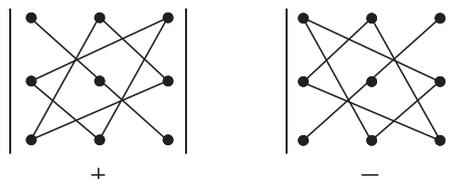
donde $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3)$, $w = (w_1, w_2, w_3)$. Así, explícitamente:

$$\begin{aligned} |A| &= (u \times v) \cdot w = \left(\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}, - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}, + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \right) (w_1, w_2, w_3) \\ &= \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} w_1 - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} w_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} w_3. \end{aligned}$$

Al desarrollar esta última expresión llegamos a la fórmula

$$|A| = u_1 v_2 w_3 + u_2 v_3 w_1 + u_3 v_1 w_2 - (u_3 v_2 w_1 + u_1 v_3 w_2 + u_2 v_1 w_3)$$

conocida como *regla de Sarrus*, que puede recordarse fácilmente con la figura siguiente:



De esta fórmula (o, mejor, de la figura) se sigue inmediatamente que el determinante de una matriz A coincide con el de la matriz traspuesta A^t que resulta de cambiar sus filas por sus columnas. En efecto, al cambiar filas por columnas, los tres sumandos negativos no se alteran y, de los positivos, uno no se altera y los otros dos se intercambian, pero el resultado es el mismo.

Por lo tanto, el determinante de una matriz puede calcularse indistintamente como el producto mixto de los vectores formados por sus filas o por sus columnas.

Las propiedades del producto mixto se traducen inmediatamente en propiedades de los determinantes:

- a) El valor absoluto del determinante de una matriz es el volumen del paralelepípedo determinado por sus filas (o del determinado por sus columnas).
- b) Una matriz tiene determinante nulo si y sólo si una de sus filas (resp. columnas) está en el plano determinado por las otras dos.
- c) El determinante de una matriz cambia de signo si se intercambian dos de sus filas (o columnas).

Conviene observar que la regla de Sarrus se puede usar también como regla mnemotécnica para el cálculo del producto vectorial, en la forma:

$$u \times v = \begin{vmatrix} e_1 & e_2 & e_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix},$$

donde

$$e_1 = (1, 0, 0), \quad e_2 = (0, 1, 0), \quad e_3 = (0, 0, 1),$$

pues al desarrollar formalmente este “determinante” se obtiene en efecto el producto vectorial.

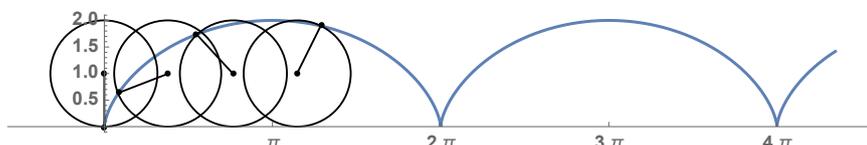
Apéndice B

Ejemplos de curvas

Recogemos en este apéndice algunos resultados sobre varias curvas que hemos tomado como ejemplo para el cálculo de longitudes, áreas y volúmenes, a fin de que los resultados sobre una misma curva no queden dispersos en las distintas secciones del libro.

B.1 La cicloide

Definición y ecuaciones La *cicloide* es la trayectoria que sigue un clavo clavado en una rueda que gira sobre una recta:



Para determinar la posición del punto en cada instante, una posibilidad es identificar los puntos del plano con \mathbb{C} y recordar que la multiplicación $z \mapsto ze^{i\alpha}$ es un giro de amplitud α alrededor de 0. Por lo tanto, un giro alrededor de un punto z_0 es $z \mapsto z_0 + (z - z_0)e^{i\alpha}$ (hacemos una traslación $z \mapsto z - z_0$ para que el centro de giro sea 0, giramos y trasladamos de nuevo $z \mapsto z + z_0$).

Si la rueda tiene radio R y gira sobre el eje real hacia la derecha, y el clavo parte de la posición 0 (luego el centro de la rueda parte de Ri), entonces, cuando el centro de la rueda se ha desplazado Rt unidades hacia la derecha, la rueda ha girado t radianes en sentido negativo, luego la posición del clavo que estaba en el punto 0 se puede obtener girándolo primero alrededor de Ri , y luego trasladándolo Rt unidades:

$$\begin{aligned} 0 \mapsto Ri - iRe^{-it} &\mapsto iR - iRe^{-it} + Rt = Rt + Ri - Ri(\cos t - i \sin t) \\ &= Rt + Ri - R \sin t - iR \cos t, \end{aligned}$$

luego la posición del punto tras un giro de t radianes es

$$r(t) = (Rt - R \operatorname{sen} t, R - R \operatorname{cos} t), \tag{B.1}$$

y ésta es una parametrización de la cicloide, que es una curva periódica con periodo (respecto a esta parametrización) 2π .

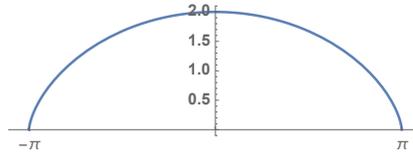
A veces conviene considerar una parametrización alternativa de modo que $t = 0$ coincida con el punto más alto, lo que supone cambiar $t \mapsto t + \pi$:

$$r(t) = (Rt + R\pi + R \operatorname{sen} t, R + R \operatorname{cos} t),$$

aunque en este caso es preferible trasladar la curva para que cumpla $x(0) = 0$, lo que supone cambiar $x \mapsto x - R\pi$, y así obtenemos:

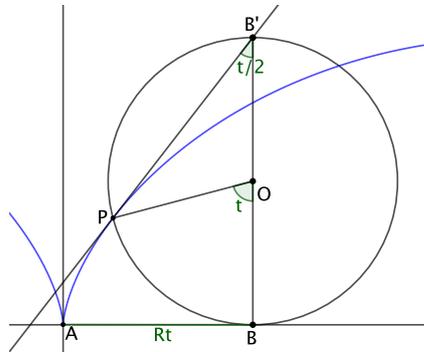
$$r(t) = R(t + \operatorname{sen} t, 1 + \operatorname{cos} t). \tag{B.2}$$

La figura siguiente muestra el arco de esta parametrización correspondiente al intervalo $-\pi \leq t \leq \pi$:



Luego usaremos una propiedad curiosa de la cicloide:

La tangente a una cicloide por un punto P es la recta que une P con el punto superior B' de la circunferencia que la genera.



En efecto, es claro que la recta PB' forma un ángulo $t/2$ con la vertical, mientras que, derivando en (B.1), el vector tangente es $r'(t) = R(1 - \operatorname{cos} t, \operatorname{sen} t)$, luego el ángulo ϕ que la tangente forma con la vertical cumple

$$\tan \phi = \frac{1 - \operatorname{cos} t}{\operatorname{sen} t} = \frac{1 - \operatorname{cos}^2(t/2) + \operatorname{sen}^2(t/2)}{2 \operatorname{sen}(t/2) \operatorname{cos}(t/2)} = \tan \frac{t}{2},$$

luego es $\phi = t/2$ y la recta tangente es precisamente PB' .

Notemos que esto implica a su vez que la normal a la cicloide es la recta PB .

Área El área de la cicloide viene dada por

$$A = \int_0^{2\pi R} y(x) dx,$$

donde $y(x)$ es, naturalmente, la función que calcula la altura $y(x)$ de la cicloide correspondiente a una abscisa x dada. No la conocemos, pero podemos eliminarla mediante el cambio de variable $x = R(t - \operatorname{sen} t)$, de modo que

$$\begin{aligned} A &= \int_0^{2\pi} R(1 - \cos t)R(1 - \cos t) dt = R^2 \int_0^{2\pi} (1 + \cos^2 t - 2 \cos t) dt \\ &= R^2 \int_0^{2\pi} \left(1 + \frac{1 + \cos 2t}{2} - 2 \cos t\right) dt = R^2 \left[\frac{3}{2}t + \frac{1}{4} \operatorname{sen} 2t - 2 \operatorname{sen} t \right]_0^{2\pi}, \end{aligned}$$

luego el área limitada por la cicloide es

$$A = 3\pi R^2.$$

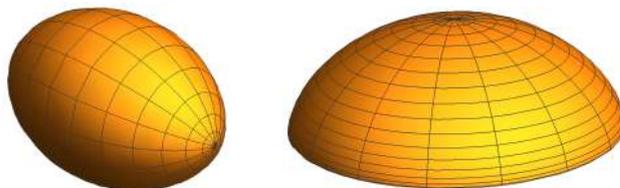
Longitud Para calcular la longitud de un arco de cicloide observamos que $r'(t) = R(1 - \cos t, \operatorname{sen} t)$, luego

$$\|r'(t)\| = R\sqrt{2 - 2 \cos t} = 2R \operatorname{sen} \frac{t}{2},$$

con lo que la longitud de un arco de cicloide es

$$L = \int_0^{2\pi} 2R \operatorname{sen} \frac{t}{2} dt = 4R \left[-\cos \frac{t}{2} \right]_0^{2\pi} = 8R.$$

Volumen del sólido de revolución Vamos a calcular el volumen limitado por la superficie que resulta de girar una cicloide respecto de uno u otro eje:



Consideramos la parametrización (B.2) y aplicamos la fórmula(2.4):

$$V = \pi \int_{-\pi R}^{\pi R} y(x)^2 dx,$$

donde $y(x)$ es la función que determina la ordenada y del punto de la cicloide de abscisa x . Como en el caso del área, eliminamos la función $y(x)$ mediante el cambio de variable $x = R(t + \operatorname{sen} t)$, con lo que

$$V = \pi \int_{-\pi}^{\pi} R^2(1 + \cos t)^2 R(1 + \cos t) dt = \pi R^3 \int_{-\pi}^{\pi} (1 + \cos t)^3 dt$$

$$\begin{aligned}
&= \pi R^3 \int_{-\pi}^{\pi} (1 + 3 \cos t + 3 \cos^2 t + \cos^3 t) dt \\
&= \pi R^3 \left[t + 3 \operatorname{sen} t + \frac{3}{2}t + \frac{3}{4} \operatorname{sen} 2t + \operatorname{sen} t - \frac{\operatorname{sen}^3 t}{3} \right]_{-\pi}^{\pi} = 5\pi^2 R^3,
\end{aligned}$$

donde hemos usado (C.1) y (C.2). Éste es el volumen de la figura de la izquierda. Para calcular el de la figura de la derecha planteamos la integral

$$\begin{aligned}
V &= \int_0^{2R} x(y) dy = - \int_{\pi}^0 R^2 (t + \operatorname{sen} t)^2 R \operatorname{sen} t dt = \\
&\pi R^3 \int_0^{\pi} (t^2 \operatorname{sen} t + 2t \operatorname{sen}^2 t + \operatorname{sen}^3 t) dt = \left(\frac{3\pi^3}{2} - \frac{8\pi}{3} \right) R^3,
\end{aligned}$$

donde hemos hecho el cambio de variable $y(t) = 1 + \cos t$ y la integral resultante la hemos calculado a partir de (C.3), (C.6) y (C.1).

Superficie de revolución El área de la superficie de revolución respecto del eje X determinada por la cicloide viene dada por la fórmula (2.8):

$$\begin{aligned}
A &= 2\pi \int_{-\pi}^{\pi} R(1 + \cos t) 2R \cos \frac{t}{2} dt = 8\pi R^2 \int_{-\pi}^{\pi} \cos^3 \frac{t}{2} dt \\
&= 16\pi R^2 \int_{-\pi/2}^{\pi/2} \cos^3 u du = 16\pi R^2 \left[\operatorname{sen} u - \frac{\operatorname{sen}^3 u}{3} \right]_{-\pi/2}^{\pi/2} = \frac{64}{3} \pi R^2,
\end{aligned}$$

donde hemos usado (C.1).

En el caso de la superficie de revolución respecto al eje Y sólo tenemos que intercambiar las coordenadas:

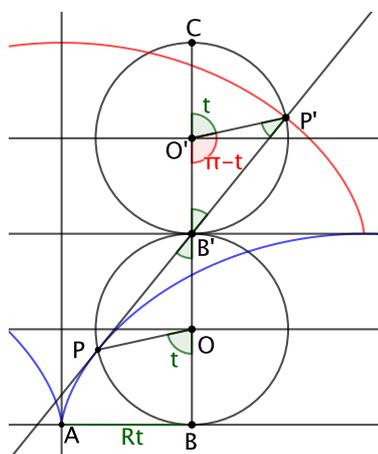
$$\begin{aligned}
A &= 2\pi \int_0^{\pi} R(t + \operatorname{sen} t) 2R \operatorname{sen} \frac{t}{2} dt = 4\pi R^2 \int_0^{\pi} (t + \operatorname{sen} t) \operatorname{sen} \frac{t}{2} dt \\
&= 4\pi R^2 \left(2 \int_0^{\pi} \frac{t}{2} \operatorname{sen} \frac{t}{2} dt + \int_0^{\pi} 2 \operatorname{sen}^2 \frac{t}{2} \cos \frac{t}{2} dt \right) \\
&= 16\pi R^2 \left(\int_0^{\pi/2} s \operatorname{sen} s ds + \int_0^{\pi} \operatorname{sen}^2 \frac{t}{2} \cos \frac{t}{2} \frac{1}{2} dt \right) \\
&= 16\pi R^2 \left(\int_0^{\pi/2} \cos s ds + \left[\frac{1}{3} \operatorname{sen}^3 \frac{t}{2} \right]_0^{\pi} \right) = \frac{64}{3} \pi R^2,
\end{aligned}$$

donde la primera integral la hemos calculado por partes. Vemos que las dos superficies tienen la misma área. ■

La evoluta de la cicloide Para terminar nuestro estudio de la cicloide demostraremos la propiedad que demostró Huygens, según la cual un péndulo de longitud L que oscila entre dos cicloides de longitud $2L$ describe una cicloide de longitud L . Equivalentemente, tenemos que probar que si ponemos fijamos un hilo de longitud L al “pico” de una cicloide de longitud $2L$ y lo enrollamos sobre ésta manteniéndolo tenso, el otro extremo describe también una cicloide de longitud $2L$.

En general, la curva obtenida de este modo a partir de cualquier curva dada se conoce como una *evoluta* de la curva (que dependerá del punto en el que fijamos el hilo y de la longitud de éste). En estos términos queremos probar que la evoluta de la cicloide de longitud $2L$ generada por un hilo de longitud L a partir de su “pico” es también una cicloide de longitud $2L$.

Para probarlo consideramos una cicloide generada por una circunferencia de radio R (con lo que su longitud es $2L = 8R$) y construimos sobre dicha circunferencia otra del mismo radio:



Hemos probado más arriba que la tangente a la cicloide por el punto P es la recta PB' . Llamamos P' al otro punto en que corta a la segunda circunferencia. Observamos que los ángulos B' miden $t/2$, luego $\widehat{B'O'P'} = \pi - t$, de donde a su vez $\widehat{CO'P'} = t$. Así pues, P' es el punto al que llega el punto superior de la circunferencia de centro $(0, 2R)$ y radio R cuando se la traslada horizontalmente Rt unidades y se la hace girar t radianes, luego P' describe también una cicloide.

Ahora basta probar que la longitud del arco AP más la del segmento $\overline{PP'}$ es precisamente $L = 4R$, pues entonces P' es precisamente la posición del otro extremo de un hilo de longitud L fijado a la cicloide por A y que está siendo enrollado en ésta (de modo que coincide con ella hasta el punto P y es tangente a partir de ahí).

En efecto, considerando cualquiera de los dos ángulos de vértice B' , tenemos que

$$\text{sen } \frac{t}{2} = \frac{R \text{ sen } t}{\overline{PB'}},$$

luego

$$\overline{PP'} = 2\overline{PB'} = \frac{2R \operatorname{sen} t}{\operatorname{sen}(t/2)} = 4R \cos \frac{t}{2}.$$

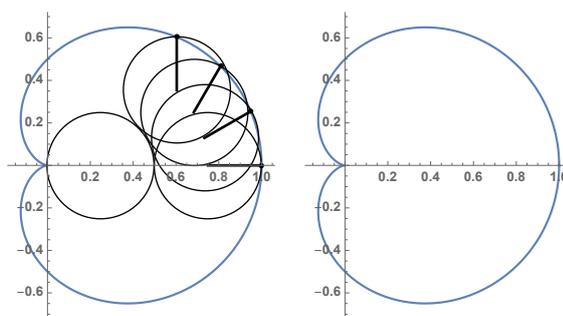
Y, por último, la longitud del arco AP es

$$\int_0^t 2R \operatorname{sen} \frac{\tau}{2} d\tau = 4R \left[-\cos \frac{\tau}{2} \right]_0^t = 4R - 4R \cos \frac{t}{2},$$

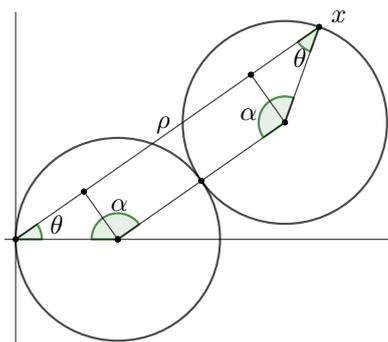
luego un hilo tenso con un extremo fijo en el punto A y el otro extremo en P tiene longitud $4R$, independientemente de t . ■

B.2 La cardioide

Definición y ecuaciones La *cardioide* es la trayectoria de un clavo clavado a una rueda que gira sobre una circunferencia. El nombre se lo dio Johann Castillon en 1741 y hace alusión a su forma de corazón:



Supongamos que la circunferencia fija tiene centro en $(a/4, 0)$ y radio¹ $a/4$, y que el clavo parte de la posición $(0, 0)$. Cuando la rueda ha girado α radianes, la situación es la que indica la figura:



¹Al llamar $a/4$ al radio de las circunferencias el parámetro a tiene una interpretación geométrica natural en términos de la propia cardioide, pues es la distancia entre los dos puntos en los que corta al eje horizontal.

El cuadrilátero tiene dos ángulos y dos lados iguales, por lo que los otros dos ángulos también tienen la misma amplitud θ . Es claro entonces que

$$\rho = \frac{a}{2} + 2\frac{a}{4} \cos \theta,$$

luego

$$\rho = \frac{a}{2}(1 + \cos \theta),$$

donde $\theta \in]-\pi, \pi[$.

Área El teorema 2.6 nos da el área limitada por una cardioide:

$$\begin{aligned} A &= \frac{a^2}{8} \int_0^{2\pi} (1 + \cos^2 \theta + 2 \cos \theta) d\theta = \frac{a^2}{8} \int_0^{2\pi} \left(\frac{3}{2} + \frac{\cos 2\theta}{2} + 2 \cos \theta \right) d\theta \\ &= \frac{a^2}{8} \left[\frac{3}{2}\theta + \frac{\sin 2\theta}{4} + 2 \sin \theta \right]_0^{2\pi} = \frac{3\pi}{8} a^2. \end{aligned}$$

Longitud La longitud de la cardioide fue calculada por Philippe de La Hire en 1708. Para calcularla usamos la fórmula (2.3) tomando como parámetro $t = \theta$, con lo que $\theta' = 1$ y

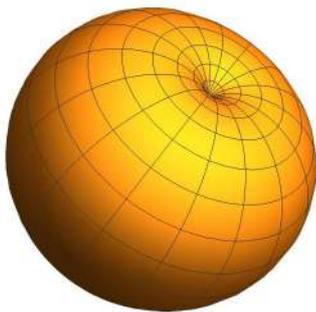
$$\rho'(\theta) = -\frac{a}{2} \sin \theta,$$

luego

$$\begin{aligned} L &= \frac{a}{2} \int_{-\pi}^{\pi} \sqrt{\sin^2 \theta + (1 + \cos \theta)^2} d\theta = \frac{a}{2} \int_{-\pi}^{\pi} \sqrt{2 + 2 \cos \theta} d\theta \\ &= a \int_{-\pi}^{\pi} \sqrt{\frac{1 + \cos \theta}{2}} d\theta = a \int_{-\pi}^{\pi} \cos \frac{\theta}{2} d\theta = 2a \int_{-\pi}^{\pi} \frac{1}{2} \cos \frac{\theta}{2} d\theta = 2a [\sin(\theta/2)]_{-\pi}^{\pi}, \end{aligned}$$

luego concluimos que la longitud de una cardioide de sección a es $L = 4a$. ■

Volumen de revolución La fórmula (2.6) nos permite calcular fácilmente el volumen que resulta de girar una cardioide sobre su eje:



$$V = \frac{2\pi}{3} \frac{a^3}{8} \int_0^{\pi} (1 + \cos \theta)^3 \sin \theta d\theta = \frac{\pi a^3}{12} \left[-\frac{(1 + \cos \theta)^4}{4} \right]_0^{\pi} = \frac{\pi a^3}{3}.$$

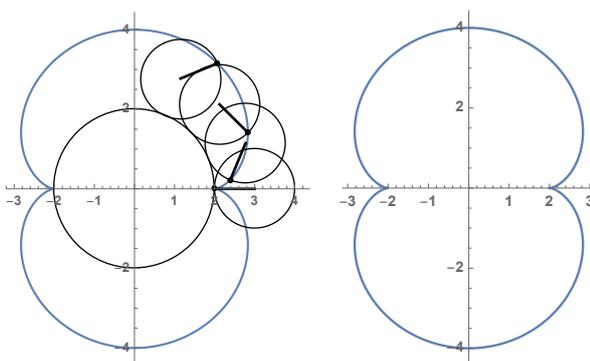
Superficie de revolución El área de la superficie de revolución la calculamos con (2.9):

$$\begin{aligned} A &= 2\pi \int_0^\pi \frac{a^2}{4} (1 + \cos \theta) \sqrt{(1 + \cos \theta)^2 + \sin^2 \theta} d\theta \\ &= \frac{\pi a^2}{2} \int_0^\pi (1 + \cos \theta) \sqrt{2(1 + \cos \theta)} d\theta = 2\pi a^2 \int_0^\pi \cos^3 \frac{\theta}{2} d\theta = \frac{8}{3} \pi a^2, \end{aligned}$$

donde el área se calcula con el cambio $t = \theta/2$ y aplicando (C.1).

B.3 La nefroide

Definición y ecuaciones La *nefroide* es la curva que describe un punto de una circunferencia de radio a que gira alrededor de otra circunferencia de radio $2a$. Aunque era conocida desde mucho antes, el nombre se lo dio el astrónomo británico Richard Anthony Proctor en 1878 y hace alusión a que tiene forma de riñón:



Si tomamos la circunferencia mayor de centro 0, cuando la circunferencia menor ha girado un ángulo θ respecto de ésta, el camino avanzado tiene longitud $2a\theta$, pero para que una rueda de radio a avance $2a\theta$ unidades, tiene que haber girado 2θ radianes. Como en el caso de la cicloide, identificamos los puntos del plano con números complejos. Si el punto que describe la nefroide parte de la posición $2a$, un giro de $2a\theta$ radianes alrededor del punto $3a$ lo transforma en

$$3a + (2a - 3a)e^{i2\theta} = 3a - ae^{i2\theta},$$

y si ahora lo giramos θ radianes alrededor de 0, pasa a ser

$$r(\theta) = e^{i\theta}(3a - ae^{i2\theta}) = 3ae^{i\theta} - ae^{i3\theta}.$$

Separando la parte real y la imaginaria obtenemos la parametrización de la nefroide:

$$r(\theta) = (3a \cos \theta - a \cos 3\theta, 3a \sin \theta - a \sin 3\theta)$$

o, alternativamente,

$$\begin{aligned} r(\theta) &= 3a(\cos \theta + i \operatorname{sen} \theta) - a(\cos \theta + i \operatorname{sen} \theta)^3 \\ &= 3a \cos \theta + i3a \operatorname{sen} \theta - a \cos^3 \theta - i3a \cos^2 \theta \operatorname{sen} \theta + 3a \cos \theta \operatorname{sen}^2 \theta + ia \operatorname{sen}^3 \theta \end{aligned}$$

luego

$$\begin{aligned} x(\theta) &= 3a \cos \theta - a \cos^3 \theta + 3a \cos \theta(1 - \cos^2 \theta) = 6a \cos \theta - 4a \cos^3 \theta, \\ y(\theta) &= 3a \operatorname{sen} \theta - 3a(1 - \operatorname{sen}^2 \theta) \operatorname{sen} \theta + a \operatorname{sen}^3 \theta = 4a \operatorname{sen}^3 \theta. \end{aligned}$$

Y así, una parametrización alternativa es:

$$r(\theta) = (6a \cos \theta - 4a \cos^3 \theta, 4a \operatorname{sen}^3 \theta).$$

Área Para calcular el área de la nefroide usamos el teorema 2.10:

$$A = \frac{1}{2} \int_0^{2\pi} (x(\theta)y'(\theta) - y(\theta)x'(\theta)) d\theta.$$

El integrando es

$$\begin{aligned} &(6a \cos \theta - 4a \cos^3 \theta)12a \operatorname{sen}^2 \theta \cos \theta + 6a \operatorname{sen} \theta(1 - 2 \cos^2 \theta)4a \operatorname{sen}^3 \theta \\ &= 72a^2 \operatorname{sen}^2 \theta \cos^2 \theta - 48a^2 \operatorname{sen}^2 \theta \cos^4 \theta + 24 \operatorname{sen}^4 \theta - 48 \operatorname{sen}^4 \theta \cos^2 \theta \\ &= 24a^2 \operatorname{sen}^4 \theta - 48a^2 \operatorname{sen}^2 \theta \cos^2 \theta + 72a^2 \operatorname{sen}^2 \theta \cos^2 \theta \\ &= 24a^2 \operatorname{sen}^4 \theta + 24a^2 \operatorname{sen}^2 \theta \cos^2 \theta = 24a^2 \operatorname{sen}^2 \theta, \end{aligned}$$

luego el área es

$$A = 12a^2 \int_0^{2\pi} \operatorname{sen}^2 \theta d\theta = 12\pi a^2,$$

donde hemos usado (C.2). ■

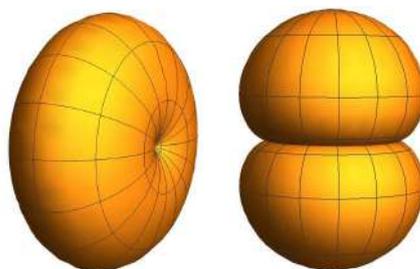
Longitud Para calcular la longitud de la nefroide calculamos:

$$\begin{aligned} \|r(x)\|^2 &= x'^2 + y'^2 = (-6a \operatorname{sen} \theta + 12a \cos^2 \theta \operatorname{sen} \theta)^2 + (12a \operatorname{sen}^2 \theta \cos \theta)^2 \\ &= 36a^2 \operatorname{sen}^2 \theta - 144a^2 \operatorname{sen}^2 \theta \cos^2 \theta + 144a^2 \cos^4 \theta \operatorname{sen}^2 \theta + 144a^2 \operatorname{sen}^4 \theta \cos^2 \theta \\ &= 36a^2 \operatorname{sen}^2 \theta - 144a^2 \operatorname{sen}^2 \theta \cos^2 \theta + 144a^2 \cos^2 \theta \operatorname{sen}^2 \theta = 36a^2 \operatorname{sen}^2 \theta, \end{aligned}$$

luego $\|r(x)\| = 6|\operatorname{sen} \theta|$. Para no tener que distinguir casos según el signo del seno de θ , calculamos la longitud de media nefroide y la multiplicamos por 2:

$$L = 12a \int_0^\pi \operatorname{sen} \theta d\theta = 24a.$$

Volumen del sólido revolución Consideramos ahora los cuerpos que resultan de girar una nefroide alrededor de los ejes:



Para calcular el volumen del primero aplicamos la fórmula (2.5):

$$\begin{aligned} V &= -\pi \int_0^\pi 16a^2 \operatorname{sen}^6 \theta (-6a \operatorname{sen} \theta + 12a \cos^2 \theta \operatorname{sen} \theta) d\theta \\ &= 96\pi a^3 \int_0^\pi (2 \operatorname{sen}^9 \theta - \operatorname{sen}^7 \theta) d\theta = \frac{1024\pi}{15} a^3, \end{aligned}$$

donde usamos la fórmula (C.9) para calcular

$$\begin{aligned} \int_0^\pi \operatorname{sen}^9 \theta d\theta &= \frac{8 \cdot 6 \cdot 4 \cdot 2}{9 \cdot 7 \cdot 5 \cdot 3} \int_0^\pi \operatorname{sen} \theta d\theta = \frac{8 \cdot 6 \cdot 4 \cdot 2}{9 \cdot 7 \cdot 5 \cdot 3} \cdot 2 = \frac{256}{315}, \\ \int_0^\pi \operatorname{sen}^7 \theta d\theta &= \frac{6 \cdot 4 \cdot 2}{7 \cdot 5 \cdot 3} \cdot 2 = \frac{32}{35}. \end{aligned}$$

Para el segundo intercambiamos las coordenadas:

$$\begin{aligned} V &= -\pi \int_{-\pi}^\pi (6a \cos \theta - 4a \cos^3 \theta)^2 12a \operatorname{sen}^2 \theta \cos \theta d\theta \\ &= 24\pi a^3 \int_{-\pi}^\pi (3 \cos^2 \theta - 5 \cos^4 \theta + 2 \cos^6 \theta) d\theta = 144\pi^2 a^3, \end{aligned}$$

donde las integrales se calculan mediante (C.9). ■

Superficie de revolución La fórmula (2.8) nos da el área de las superficies de revolución generadas por la nefroide. Notemos que al calcular su longitud hemos hallado $\sqrt{x'^2 + y'^2} = 6a \operatorname{sen} \theta$. Por lo tanto, para la superficie de revolución respecto del eje X tenemos:

$$A = 2\pi \int_0^\pi 4a \operatorname{sen}^3 \theta \cdot 6a \operatorname{sen} \theta d\theta = 48\pi a^2 \int_0^\pi \operatorname{sen}^4 \theta d\theta = 18\pi^2 a^2,$$

donde la integral se calcula mediante (C.9).

B.4 La tractriz

Definición y ecuaciones Consideremos un cuerpo puntual situado en $(l, 0)$ atado a una cuerda de longitud l con su otro extremo en $(0, 0)$. Estiramos de la cuerda de modo que su extremo suba por el eje Y . La trayectoria del cuerpo arrastrado por la cuerda recibe el nombre de *tractriz*.

En realidad, se considera que la tractriz completa consta también de la parte simétrica respecto del eje X que resulta de estirar también la cuerda descendiendo por el eje Y . Vamos a parametrizarla.

Sea (x, y) un punto arbitrario de la tractriz, sea $(0, lt)$ el punto donde se encuentra el otro extremo de la cuerda y sea θ el ángulo que muestra la figura, de modo que $0 < \theta \leq \pi/2$ cuando $t \leq 0$ y $\pi/2 \leq \theta < \pi$ cuando $t > 0$. Entonces, teniendo en cuenta que $\cos \theta$ es negativo si t es positivo y viceversa,

$$(x, y) = (l \operatorname{sen} \theta, lt + l \cos \theta).$$

Pero t y θ no son independientes. La clave para relacionarlos es que un cuerpo estirado por una cuerda se mueve en la dirección de la cuerda, luego ésta ha de ser tangente a la trayectoria.

Por lo tanto, si consideramos a x, y, θ como funciones del parámetro t , el vector tangente $(x'(t), y'(t))$ tiene que tener la dirección de la cuerda, lo que se traduce en que

$$\frac{y'(t)}{x'(t)} = \frac{\cos \theta(t)}{\operatorname{sen} \theta(t)}.$$

Explícitamente:

$$\frac{l - l \operatorname{sen} \theta \theta'}{l \cos \theta \theta'} = \frac{\cos \theta}{\operatorname{sen} \theta}.$$

Despejando: $\operatorname{sen} \theta - \operatorname{sen}^2 \theta \theta' = \cos^2 \theta \theta'$, luego $\theta' = \operatorname{sen} \theta$. Esto es una ecuación diferencial, a la que podemos añadir la condición inicial $\theta(0) = \pi/2$. Separamos variables e integramos:

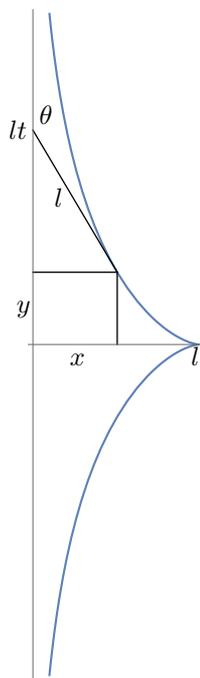
$$\int \frac{d\theta}{\operatorname{sen} \theta} = \int dt.$$

Para la primera integral aplicamos el cambio de variables de Weierstrass $u = \tan(\theta/2)$. Entonces $\theta = 2 \arctan u$, con lo que

$$d\theta = \frac{2}{1 + u^2} du,$$

mientras que

$$\operatorname{sen} \theta = 2 \operatorname{sen} \frac{\theta}{2} \cos \frac{\theta}{2} = 2 \tan \frac{\theta}{2} \cos^2 \frac{\theta}{2} = \frac{2 \tan(\theta/2)}{1 + \tan^2(\theta/2)} = \frac{2u}{1 + u^2}.$$



Así pues:

$$\int \frac{d\theta}{\operatorname{sen} \theta} = \int \frac{1+u^2}{2u} \frac{2}{1+u^2} du = \int \frac{1}{u} du = \log u = \log \tan \frac{\theta}{2}.$$

Por lo tanto, al integrar la ecuación diferencial nos queda:

$$\log \tan \frac{\theta}{2} = t + c,$$

y la condición inicial nos permite calcular $c = 0$. En definitiva:

$$t = \log \tan \frac{\theta}{2},$$

luego una parametrización de la tractriz en función de θ es

$$r(\theta) = (l \operatorname{sen} \theta, l \log \tan \frac{\theta}{2} + l \cos \theta), \quad 0 < \theta < \pi. \quad (\text{B.3})$$

Si queremos usar t como parámetro tenemos que despejar:

$$y = \tan \frac{\theta}{2} = e^t,$$

luego

$$\begin{aligned} \operatorname{sen} \theta &= \frac{2u}{1+u^2} = \frac{2e^t}{1+e^{2t}} = \frac{2}{e^t + e^{-t}} = \frac{1}{\cosh t}, \\ \cos^2 \theta &= 1 - \frac{1}{\cosh^2 t} = \frac{\cosh^2 t - 1}{\cosh^2 t} = \frac{\operatorname{senh}^2 t}{\cosh^2 t}, \end{aligned}$$

luego, teniendo en cuenta que $\cos \theta$ es negativo,

$$\cos \theta = -\tanh t.$$

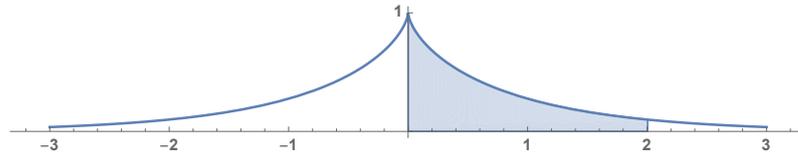
En total queda:

$$r(t) = \left(\frac{l}{\cosh t}, lt - l \tanh t \right). \quad (\text{B.4})$$

Alternativamente, si intercambiamos las funciones coordenadas tenemos la tractriz con su asíntota en el eje horizontal:

$$r(t) = (x(t), y(t)) = \left(lt - l \tanh t, \frac{l}{\cosh t} \right). \quad (\text{B.5})$$

Área Considerando la parametrización (B.5), vamos a calcular el área comprendida entre la tractriz y su asíntota, más concretamente entre $x = 0$ y $x = x(t_0)$:



En principio podemos expresarla en términos de la función desconocida $y(x)$ cuya gráfica es la tractriz, pero inmediatamente la eliminamos con el cambio de variable $x = x(t)$:

$$\begin{aligned} A_{t_0} &= \int_0^{x(t_0)} y(x) dx = \int_0^{t_0} y(t)x'(t) dt = l^2 \int_0^{t_0} \frac{\tanh^2 t}{\cosh t} dt = l^2 \int_0^{t_0} \frac{\sinh^2 t}{\cosh^3 t} dt \\ &= -l^2 \frac{\tanh t_0}{2 \cosh t_0} + \frac{l^2}{2} \arctan \sinh t_0 \end{aligned}$$

Lo más interesante es que

$$\lim_{t_0 \rightarrow +\infty} A_{t_0} = \frac{\pi l^2}{4},$$

lo que significa que, aunque la tractriz sea infinita, concretamente, incluyendo también la parte negativa, vemos que el área total limitada por la tractriz es

$$A = \frac{\pi l^2}{2}. \quad \blacksquare$$

Longitud Para calcular la longitud de un arco de tractriz usamos la parametrización en términos de θ , cuya derivada es:

$$r'(\theta) = \left(l \cos \theta, l \frac{\cos^2 \theta}{\sin \theta} \right), \quad \|r'(\theta)\| = l \frac{|\cos \theta|}{\sin \theta}.$$

Vemos así que la tractriz es regular en salvo en $\theta = \pi/2$, y la longitud de un arco de tractriz es

$$s(\theta) = -l \int_{\pi/2}^{\theta} \frac{\cos t}{\sin t} dt = -l [\log \sin t]_{\pi/2}^{\theta} = -l \log \sin \theta \quad \blacksquare$$

La pseudoesfera La superficie que resulta del giro de una tractriz respecto de su asíntota se llama *pseudoesfera*. Para calcular su volumen con la fórmula (2.4) consideramos la parametrización (B.5).

Para aplicar la fórmula del volumen necesitamos considerar la función $y(x) = y(t(x))$, pero en la integral podemos hacer el cambio de variable $x = x(t)$, con lo que

$$dx = x'(t) dt = l \tanh^2 t dt$$

para volver a una expresión en función de t :

$$\begin{aligned} V_{t_0} &= \pi \int_0^{x_0} y(x)^2 dx = \pi \int_0^{t_0} y(t)^2 l \tanh^2 t dt \\ &= l^3 \pi \int_0^{t_0} \frac{1}{\cosh^2 t} \tanh^2 t dt = \frac{l^3 \pi}{3} \tanh^3 t_0. \end{aligned}$$



Éste es el volumen de una sección de pseudoesfera desde su base, correspondiente a $t = 0$, hasta la altura determinada por el valor t_0 del parámetro. Como en el caso del área de la tractriz, el volumen de la pseudoesfera completa resulta ser finito:

$$\lim_{t_0 \rightarrow +\infty} V_{t_0} = \lim_{t_0 \rightarrow +\infty} \frac{l^3 \pi}{3} \tanh^3 t_0 = \frac{1}{3} \pi l^3,$$

al igual que el volumen de la pseudoesfera completa, que resulta ser

$$V = \frac{2}{3} \pi l^3,$$

la mitad del volumen de una esfera de radio l . ■

El área de la pseudoesfera Recordemos que una parametrización de la pseudoesfera con su asíntota en el eje de abscisas viene dada por

$$x(t) = lt - l \tanh t, \quad y(t) = \frac{l}{\cosh t},$$

de modo que

$$x'(t) = l \tanh^2 t, \quad y'(t) = -\frac{l}{\cosh^2 t} \sinh t = -\frac{l \tanh t}{\cosh t}.$$

Para calcular el área de un tronco de pseudoesfera, desde $x = 0$ hasta $x = x_0$, necesitamos la función $y(t(x))$, pero tras aplicar la fórmula (2.7) deshacemos el cambio de variable:

$$\begin{aligned} A_{t_0} &= 2\pi \int_0^{x_0} y(x) \sqrt{1 + y'(x)^2} dx = 2\pi \int_0^{t_0} y(t) \sqrt{1 + \frac{y'^2(t)}{x'^2(t)}} x'(t) dt = \\ &= 2\pi l^2 \int_0^{t_0} \frac{1}{\cosh t} \sqrt{1 + \frac{1}{\sinh^2 t}} \tanh^2 t dt = 2\pi l^2 \int_0^{t_0} \frac{1}{\cosh t} \sqrt{\frac{\cosh^2 t}{\sinh^2 t}} \tanh^2 t dt \\ &= 2\pi l^2 \int_0^{t_0} \frac{\sinh t}{\cosh^2 t} dt = -2\pi l^2 \left[\frac{1}{\cosh t} \right]_0^{t_0} = 2\pi l^2 - \frac{2\pi l^2}{\cosh t_0}. \end{aligned}$$

Ahora vemos que

$$\lim_{t_0 \rightarrow +\infty} A_{t_0} = \lim_{t_0 \rightarrow +\infty} 2\pi l^2 - \frac{2\pi l^2}{\cosh t_0} = 2\pi l^2,$$

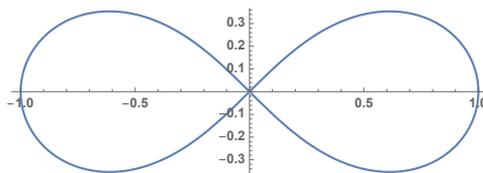
y esta integral puede verse como el área de media pseudoesfera, con lo que el área de la pseudoesfera completa resulta ser $A = 4\pi l^2$.

Vemos así que una pseudoesfera tiene la misma área y la mitad del volumen de la esfera de radio l .

B.5 La lemniscata de Bernoulli

Definición y ecuaciones La *lemniscata de Bernoulli* es la curva dada por la ecuación cartesiana

$$(x^2 + y^2)^2 = a^2(x^2 - y^2), \quad (\text{B.6})$$



La sustitución $x = \rho \cos \theta$, $y = \rho \operatorname{sen} \theta$ nos da

$$\rho^4 = a^2 \rho^2 \cos 2\theta,$$

que a su vez se simplifica a la expresión de la lemniscata en coordenadas polares:

$$\rho = a\sqrt{\cos 2\theta}.$$

Esta ecuación muestra claramente que la curva tiene, en efecto, forma de 8, y que el parámetro a es la semiamplitud, es decir, la mitad de la distancia entre sus puntos más alejados. Un bucle corresponde a los ángulos $-\pi/4 \leq \theta \leq \pi/4$ y el otro a $3\pi/4 \leq \theta \leq 5\pi/4$. Para otros argumentos el coseno es negativo y no hay ningún punto correspondiente en la lemniscata.

Área El teorema 2.6 nos da que el área de la lemniscata es

$$A = \int_{-\pi/4}^{\pi/4} a^2 \cos 2\theta \, d\theta = \left[\frac{a^2}{2} \operatorname{sen} 2\theta \right]_{-\pi/4}^{\pi/4} = a^2.$$

Longitud La fórmula (2.3) nos da que la longitud de una lemniscata, expresada como cuatro veces la longitud de un cuarto de la curva, es

$$L = 4 \int_0^{\pi/4} \sqrt{\rho'^2 + \rho^2} \, d\theta.$$

Tenemos que

$$\rho' = -\frac{a \operatorname{sen} 2\theta}{\sqrt{\cos 2\theta}},$$

luego

$$\rho'^2 + \rho^2 = \frac{a^2 \operatorname{sen}^2 2\theta}{\cos 2\theta} + a^2 \cos 2\theta = \frac{a^2}{\cos 2\theta},$$

y así

$$L = 4a \int_0^{\pi/4} \frac{1}{\sqrt{\cos 2\theta}} \, d\theta = 2a \int_0^{\pi/2} \frac{1}{\sqrt{\cos \theta}} \, d\theta.$$

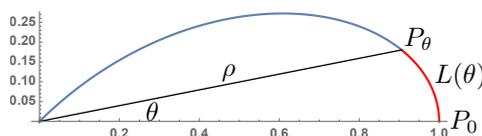
El teorema 6.4 reduce la integral a una integral elíptica:

$$L = 2\sqrt{2}a K(1/2) \approx 5.24412a.$$

Un poco más en general, hemos obtenido que la función

$$L(\theta) = 4a \int_0^\theta \frac{1}{\sqrt{\cos 2t}} dt, \quad 0 < \theta \leq \pi/4,$$

proporciona la longitud del arco de lemniscata comprendido entre $P_0 = (a, 0)$ y el punto P_θ de argumento θ , pero tiene interés la expresión para la longitud del arco comprendido entre $(0, 0)$ y P_θ expresada en función de ρ .



Para ello observamos que, en virtud de (B.6), las coordenadas (x, y) del punto del primer cuadrante de coordenadas (ρ, θ) cumple

$$x^2 + y^2 = \rho^2, \quad a^2(x^2 - y^2) = \rho^4.$$

Sumando y restando a la segunda ecuación la primera multiplicada por a^2 queda

$$2a^2x^2 = \rho^4 + a^2\rho^2, \quad -2a^2y^2 = \rho^4 - a^2\rho^2,$$

luego

$$(x, y) = \left(\sqrt{\frac{1}{2a^2}\rho^2(a^2 + \rho^2)}, \sqrt{\frac{1}{2a^2}\rho^2(a^2 - \rho^2)} \right),$$

y la fórmula para la longitud a partir de esta parametrización, tras un cálculo rutinario, nos da que

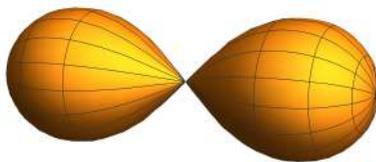
$$L(\rho) = \int_0^\rho \sqrt{x'^2(r) + y'^2(r)} dr = a^2 \int_0^\rho \frac{1}{\sqrt{a^4 - r^4}} dr = a \int_0^{\rho/a} \frac{1}{\sqrt{1 - r^4}} dr.$$

En particular, la longitud de la lemniscata completa es

$$L = 4a \int_0^1 \frac{1}{\sqrt{1 - r^4}} dr.$$

Volumen del sólido de revolución Para calcular el volumen del sólido de revolución determinado por una lemniscata usamos la fórmula (2.4), junto con la expresión (B.10), que para la lemniscata $c = d = a/\sqrt{2}$ se reduce a:

$$y = \sqrt{\sqrt{a^4/4 + 2a^2x^2} - x^2 - a^2/2}.$$



Así, usando (C.8):

$$\begin{aligned}
 V &= 2\pi \int_0^a (\sqrt{a^4/4 + 2a^2x^2} - x^2 - a^2/2) dx = \\
 &= 2\pi \left(\frac{a^2}{2} \int_0^a \sqrt{1 + (2\sqrt{2}x/a)^2} dx - a^3/3 - a^3/2 \right) = \\
 &= \frac{1}{2\sqrt{2}} \pi a^3 \int_0^{2\sqrt{2}} \sqrt{1+t^2} dt - \frac{5}{3} \pi a^3 = \\
 &= \frac{1}{2\sqrt{2}} \pi a^3 \left[\frac{t}{2} \sqrt{1+t^2} + \frac{1}{2} \operatorname{arsenh} t \right]_0^{2\sqrt{2}} - \frac{5}{3} \pi a^3 \\
 &= \frac{1}{2\sqrt{2}} \pi a^3 \left(3\sqrt{2} + \frac{1}{2} \operatorname{arsenh} 2\sqrt{2} \right) - \frac{5}{3} \pi a^3 \\
 &= \left(\frac{\sqrt{2}}{8} \operatorname{arsenh} 2\sqrt{2} - \frac{1}{6} \right) \pi a^3 \approx 0.145 \pi a^3.
 \end{aligned}$$

■

Superficie de revolución El área de la superficie de revolución determinada por la lemniscata se calcula fácilmente con la fórmula (2.9):

$$\begin{aligned}
 A &= 4\pi \int_0^{\pi/4} \rho \operatorname{sen} \theta \sqrt{\rho^2 + \rho'^2} d\theta = 4\pi \int_0^{\pi/4} a \sqrt{\cos 2\theta} \operatorname{sen} \theta \frac{a}{\sqrt{\cos 2\theta}} d\theta \\
 &= 4\pi a^2 \int_0^{\pi/4} \operatorname{sen} \theta d\theta = 4\pi a^2 [-\cos \theta]_0^{\pi/4} = 2\pi(2 - \sqrt{2})a^2,
 \end{aligned}$$

donde hemos aprovechado el cálculo que ya habíamos hecho para obtener la longitud de la lemniscata. ■

B.6 Los óvalos de Cassini

Definición y ecuaciones Mientras una elipse está formada por los puntos del plano tales que la suma de las distancias a dos focos es constante, un *óvalo de Cassini* está formado por los puntos del plano tales que el producto de las distancias a dos focos es constante.

Si tomamos un sistema de referencia con los focos en los puntos $(\pm c, 0)$ y llamamos d^2 al valor constante del producto, los puntos (x, y) del óvalo son los que cumplen

$$((x+c)^2 + y^2)((x-c)^2 + y^2) = d^4,$$

que equivale a

$$(x^2 + y^2 + c^2 + 2cx)(x^2 + y^2 + c^2 - 2cx) = d^4,$$

o también a

$$(x^2 + y^2 + c^2)^2 - 4c^2x^2 = d^4. \quad (\text{B.7})$$

Observemos que la homotecia $(x, y) \mapsto (kx, ky)$ transforma el óvalo que cumple la ecuación anterior en el conjunto de puntos que cumplen la ecuación

$$((x/k)^2 + (y/k)^2 + c^2)^2 - 4c^2(x/k)^2 = d^4,$$

que, multiplicando por k^4 , equivale a

$$(x^2 + y^2 + (kc)^2)^2 - 4(kc)^2x^2 = (kd)^4,$$

luego se trata del óvalo determinado por kc y kd . Por lo tanto, si definimos la *excentricidad* de un óvalo de Cassini como $e = c/d$, hemos probado que dos óvalos homotéticos tienen la misma excentricidad, y el recíproco también es cierto: si los óvalos determinados por (c, d) y (c', d') tienen la misma excentricidad, entonces $c/d = c'/d'$, luego $c/c' = d/d' = k$, y uno se transforma en otro mediante una homotecia de razón k .

Desarrollando el cuadrado en (B.7) queda:

$$(x^2 + y^2)^2 + c^4 - 2c^2x^2 + 2c^2y^2 - 4c^2x^2 = d^4$$

luego

$$(x^2 + y^2)^2 - 2c^2(x^2 - y^2) = d^4 - c^4, \quad (\text{B.8})$$

y de aquí obtenemos una ecuación sencilla para las coordenadas polares de los puntos de un óvalo de Cassini:

$$\rho^4 - 2c^2\rho^2 \cos 2\theta = d^4 - c^4. \quad (\text{B.9})$$

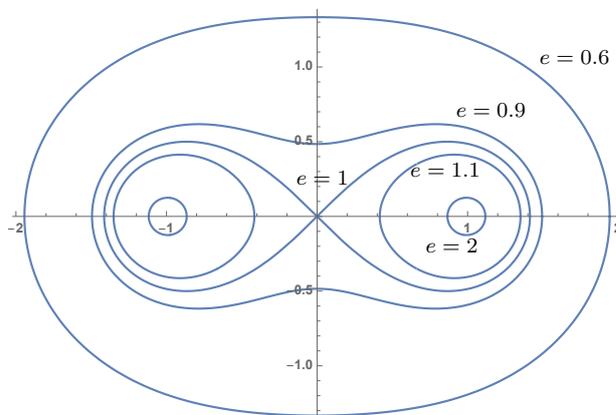
Resolviendo la ecuación de segundo grado en ρ^2 queda

$$\rho^2 = c^2(\cos 2\theta \pm e^2 \sqrt{1 - e^4 \sin^2 2\theta}),$$

o también

$$\rho = c \sqrt{\cos 2\theta \pm e^2 \sqrt{1 - e^4 \sin^2 2\theta}}.$$

Hemos visto en la introducción que los óvalos de Cassini de excentricidad pequeña se parecen a elipses, pero para excentricidades mayores el parecido desaparece. La figura siguiente muestra los óvalos correspondientes a una distancia focal $c = 1$ y varias excentricidades:



Veamos qué podemos decir en general sobre la forma de un óvalo de Cassini en función de su excentricidad. Partimos de la ecuación (B.7), en la que podemos despejar:

$$y = \pm \sqrt{\sqrt{d^4 + 4c^2x^2} - x^2 - c^2}. \quad (\text{B.10})$$

Esta expresión muestra que los óvalos son simétricos respecto de los ejes de coordenadas. El radicando interior es siempre positivo, mientras que el exterior estará definido si y sólo si

$$x^2 + c^2 \leq \sqrt{d^4 + 4c^2x^2},$$

que equivale a $x^4 + 2x^2x^2 + c^4 \leq d^4 + 4c^2x^2$, y a $x^4 - 2c^2x^2 + c^4 \leq d^4$, o también a $(x^2 - c^2)^2 \leq d^4$, o a $|x^2 - c^2| \leq d^2$, que a su vez equivale a

$$c^2 - d^2 \leq x^2 \leq c^2 + d^2. \quad (\text{B.11})$$

A partir de aquí tenemos que distinguir casos. Ante todo, es fácil ver que si $e = 0$, es decir, si $c = 0$, el óvalo de Cassini correspondiente es la circunferencia de radio d .

Si $0 < e < 1$, es decir, si $0 < c < d$, la primera de las dos desigualdades anteriores se cumple trivialmente, luego el óvalo de Cassini corta a cada recta vertical de abscisa x si y sólo si $-a \leq x \leq a$, donde $a = \sqrt{d^2 + c^2}$ es el *semieje mayor* del óvalo.

Además, se cumple $y = 0$ si y sólo si $x^2 + c^2 = \sqrt{d^4 + 4c^2x^2}$, que equivale a $|x^2 - c^2| = d^2$, lo que a su vez equivale a $x^2 = c^2 \pm d^2$, pero el signo negativo es imposible, luego $x = \pm a$. Vemos así que el óvalo puede descomponerse en dos arcos que unen los puntos $(\pm a, 0)$, uno con ordenadas estrictamente positivas y otro con ordenadas estrictamente negativas, lo que equivale a que forman una curva cerrada que rodea los dos focos, como las que muestra la figura para $e = 0.6$ y $e = 0.9$. El *semieje menor* es $b = \sqrt{d^2 - c^2}$.

Si $e = 1$, es decir, si $c = d$, el razonamiento anterior vale igualmente con la única diferencia de que ahora $y = 0$ se cumple si y sólo si $x = \pm a$ o si $x = 0$, y las ecuaciones (B.8) o (B.9) se reducen claramente a las de la lemniscata de Bernoulli.

Finalmente, si $e > 1$, es decir, si $0 < d < c$, para que el óvalo contenga puntos con abscisa x , las dos condiciones de (B.11) son relevantes, luego si llamamos $a^* = \sqrt{c^2 - d^2}$, resulta que el óvalo está definido en $[-a, -a^*]$ y en $[a^*, a]$, y además se cumple $y = 0$ si y sólo si $x = \pm a, \pm a^*$, y esto se traduce en que en realidad el óvalo está formado por dos curvas cerradas que rodean a cada uno de los focos. En la figura vemos que en el caso $e = 1.1$ tienen forma de huevo, por lo que son óvalos propiamente dichos, mientras que a medida que la excentricidad va aumentando se parecen cada vez más a dos circunferencias.

Pero todavía queda un aspecto que precisar sobre la forma de un óvalo de Cassini:

¿Para qué excentricidades $e > 1$ tienen forma de cacahuete, como el caso $e = 0.9$ que muestra la figura, y para cuáles se parecen a elipses, como en el caso $e = 0.6$?

Para responder a esta pregunta derivamos la función $f(x) = y^2(x)$ dada por la relación (B.10):

$$f'(x) = \frac{4c^2x}{\sqrt{d^4 + 4c^2x^2}} - 2x = 2x \left(\frac{2c^2}{\sqrt{d^4 + 4c^2x^2}} - 1 \right).$$

El óvalo tendrá forma de cacahuete si y sólo si $y(x)$ —luego también $f(x)$ — es decreciente a la izquierda de 0 y creciente a la derecha, es decir, en virtud del teorema 1.14, si $f'(x) < 0$ para $x < 0$ suficientemente pequeño y $f'(x) > 0$ para $x > 0$ suficientemente pequeño. Esto equivale a que

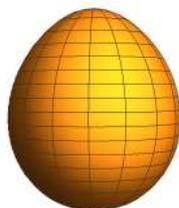
$$\frac{2c^2}{\sqrt{d^4 + 4c^2x^2}} - 1 > 0$$

o también a que $4c^4 > d^4 + 4c^2x^2$. Esto tiene que cumplirse para todo x suficientemente pequeño, lo cual sucederá si y sólo si $4c^4 > d^4$, es decir, si y sólo si $e^4 > 1/4$, si y sólo si $e > 1/\sqrt{2}$. En caso contrario, es decir, si $e \leq 1/\sqrt{2}$, se cumple que $f'(x) < 0$ para todo $x > 0$ y $f'(x) > 0$ para todo $x < 0$, luego la curva crece desde $-a$ hasta 0 y decrece desde 0 hasta a . En resumen, el aspecto de los óvalos de Cassini depende de su excentricidad según se indica en la tabla siguiente:

Excentricidad	Aspecto
$e = 0$	circunferencia
$0 < e \leq \sqrt{2}/2$	parecida a una elipse de semiejes $\sqrt{d^2 \pm c^2}$
$\sqrt{2}/2 < e < 1$	cacahuete de semiejes $\sqrt{c^2 \pm d^2}$
$e = 1$	lemniscata de semieje $a = \sqrt{2}c$
$e > 1$	óvalos comprendidos entre las abscisas $\pm\sqrt{c^2 \pm d^2}$

Ejercicio: Calcular los puntos de altura y máxima de un óvalo de Cassini.

Ejemplo Vamos a calcular el volumen del huevo que resulta de girar (la mitad de) el óvalo de Cassini de excentricidad $e = 1.05$ y distancia focal $c = 1$:



La ecuación del óvalo es (B.10), con $d = 1/1.05$, luego, según la fórmula (2.4),

$$V = \pi \int_{a^*}^a (\sqrt{d^4 + 4x^2} - x^2 - 1) dx,$$

donde $a^* = \sqrt{1 - d^2}$, $a = \sqrt{1 + d^2}$. Tenemos que

$$V = d^2 \pi \int_{a^*}^a \sqrt{1 + (2x/d^2)^2} dx - \pi \int_{a^*}^a (x^2 + 1) dx.$$

La primera integral se reduce a (C.8) mediante el cambio de variable $t = 2x/d^2$. Omitimos los cálculos rutinarios. El resultado es $V \approx 0.451$. ■

Apéndice C

Cálculo de primitivas

C.1 Integrales inmediatas

Se llaman *integrales inmediatas* las que pueden calcularse sin más que aplicar las reglas de derivación a la inversa, a lo sumo con manipulaciones sencillas del integrando (aunque sencillas no quiere decir necesariamente que sea fácil caer en su conveniencia).

La tabla C.1 contiene las reglas de derivación inversas. (Véase la página 2.2 para su justificación.) Una aplicación básica consiste en identificar una posible regla que pueda adecuarse a una integral dada, identificar la función $f(x)$ adecuada, comprobar que, en efecto, el integrando se ajusta a la regla, y aplicarla.

Ejemplo

$$\int 3x^2 \operatorname{sen} x^3 dx = -\cos x^3 + c.$$

Esto resulta de identificar la cuarta regla de la tabla como de posible aplicación, tomar $f(x) = x^3$ (porque en dicha regla $f(x)$ es el argumento del seno), comprobar que delante tenemos ciertamente $f'(x) = 3x^2$ y aplicar la regla. ■

Entre las manipulaciones más básicas que podemos hacer a una integral para ajustarla a una (o varias) de las reglas de integración se encuentra el aprovechar la linealidad de la integral, es decir, el hecho de si F_1 y F_2 son primitivas de f_1 y f_2 entonces $aF_1 + bF_2$ es una primitiva de $af_1 + bf_2$ se puede enunciar alternativamente como que

$$\int (af_1(x) + bf_2(x)) dx = a \int f_1(x) dx + b \int f_2(x) dx.$$

Ejemplo

$$\begin{aligned} \int (10x \operatorname{sen} x^2 + 6x^2 \cos x^3) dx &= 5 \int 2x \operatorname{sen} x^2 dx + 2 \int 3x^2 \cos x^3 dx \\ &= -4 \cos x^2 + 2 \operatorname{sen} x^3 + c. \end{aligned} \quad \blacksquare$$

Tabla C.1: Integrales inmediatas

$\int f(x)^n f'(x) dx$	$= \frac{f(x)^{n+1}}{n+1} + c, \quad n \neq -1$
$\int \frac{f'(x)}{f(x)} dx$	$= \log f(x) + c$
$\int f'(x) a^{f(x)} dx$	$= \frac{1}{\log a} a^{f(x)} + c$
$\int f'(x) \operatorname{sen} f(x) dx$	$= -\cos f(x) + c$
$\int f'(x) \operatorname{cos} f(x) dx$	$= \operatorname{sen} f(x) + c$
$\int \frac{f'(x)}{\sqrt{1-f(x)^2}} dx$	$= \operatorname{arcsen} f(x) + c$
$\int \frac{-f'(x)}{\sqrt{1-f(x)^2}} dx$	$= \operatorname{arccos} f(x) + c$
$\int \frac{f'(x)}{1+f(x)^2} dx$	$= \operatorname{arctan} f(x) + c$
$\int f'(x) \operatorname{senh} f(x) dx$	$= \operatorname{cosh} f(x) + c$
$\int f'(x) \operatorname{cosh} f(x) dx$	$= \operatorname{senh} f(x) + c$
$\int \frac{f'(x)}{\sqrt{f(x)^2+1}} dx$	$= \operatorname{arsenh} f(x) + c$
$\int \frac{f'(x)}{\sqrt{f(x)^2-1}} dx$	$= \operatorname{arcosh} f(x) + c$
$\int \frac{f'(x)}{1-f(x)^2} dx$	$= \operatorname{artanh} f(x) + c$

Notemos que en el ejemplo anterior hemos dejado oportunamente los números adecuados dentro de las integrales para que se pudieran aplicar las reglas. Un poco más en general, es posible “ajustar constantes” multiplicando y dividiendo el integrando por cantidades adecuadas.

Ejemplo Si queremos calcular

$$\int \frac{7}{3\sqrt{x}} e^{5\sqrt{x}} dx,$$

podemos tratar de aplicar la tercera regla con $f(x) = 5\sqrt{x}$, pero entonces

$$f'(x) = \frac{5}{2\sqrt{x}},$$

que no es exactamente lo que tenemos.

Ahora bien, podemos hacer lo siguiente:

$$\int \frac{7}{3\sqrt{x}} e^{5\sqrt{x}} dx = \frac{7}{3} \int \frac{1}{\sqrt{x}} e^{5\sqrt{x}} dx = \frac{7}{3} \cdot \frac{2}{5} \int \frac{5}{2\sqrt{x}} e^{5\sqrt{x}} dx = \frac{14}{15} e^{5\sqrt{x}} + c,$$

donde primer hemos sacado las constantes indeseadas por linealidad y luego hemos multiplicado y dividido el integrando por 5 y por 2, dejando dentro únicamente las constantes que nos interesan para ajustar la derivada, y así ya hemos podido aplicar la regla. ■

Familiarizarse con este tipo de ajustes es lo más elemental que tiene que conseguir el lector que quiera conseguir cierto dominio del cálculo de primitivas. De aquí en adelante los ajustes de este tipo los haremos sin más comentario.

Pero las reglas de integración inmediata permiten calcular muchas más integrales en función del ingenio de cada cual. Veamos algunos casos:

Ejemplos

$$\int \tan x dx = \int \frac{\sen x}{\cos x} dx = -\log |\cos x| + c. \quad \blacksquare$$

$$\int \sen^3 x dx = -\cos x + \frac{\cos^3 x}{3} + c \quad \int \cos^3 x dx = \sen x - \frac{\sen^3 x}{3} + c. \quad (\text{C.1})$$

En efecto:

$$\begin{aligned} \int \sen^3 x dx &= \int (1 - \cos^2 x) \sen x dx = \int \sen x dx - \int \cos^2 x \sen x dx \\ &= -\cos x + \frac{\cos^3 x}{3} + c, \end{aligned}$$

e igualmente se calcula la segunda integral. ■

Al tratar con funciones trigonométricas suelen ser útiles las relaciones del ángulo doble y mitad:

$$\begin{aligned} \sen 2x &= 2 \sen x \cos x, & \cos 2x &= \cos^2 x - \sen^2 x, \\ \sen \frac{x}{2} &= \sqrt{\frac{1 - \cos x}{2}}, & \cos \frac{x}{2} &= \sqrt{\frac{1 + \cos x}{2}}. \end{aligned}$$

Por ejemplo, las integrales siguientes nos aparecerán a menudo:

Ejemplo Las integrales siguientes nos aparecerán a menudo:

$$\int \sen^2 x dx = \frac{x}{2} - \frac{1}{4} \sen 2x + c, \quad \int \cos^2 x dx = \frac{x}{2} + \frac{1}{4} \sen 2x + c. \quad (\text{C.2})$$

En particular

$$\int_0^\pi \sen^2 x dx = \int_{-\pi/2}^{\pi/2} \cos^2 x dx = \frac{\pi}{2}.$$

En efecto, por las fórmulas para el ángulo mitad aplicadas a $2x$:

$$\int \operatorname{sen}^2 x \, dx = \int \frac{1 - \cos 2x}{2} \, dx = \frac{x}{2} - \frac{1}{4} \int 2 \cos 2x \, dx = \frac{x}{2} - \frac{1}{4} \operatorname{sen} 2x + c,$$

e igualmente se razona con la segunda integral. ■

C.2 Funciones racionales

Las integrales de polinomios son todas inmediatas:

$$\int (a_n x^n + \cdots + a_1 x + a_0) \, dx = a_n \frac{x^{n+1}}{n+1} + \cdots + a_1 \frac{x^2}{2} + a_0 x + c,$$

y ahora vamos a ver un método para integrar cualquier cociente de polinomios, es decir, para resolver cualquier integral de la forma

$$\int \frac{P(x)}{Q(x)} \, dx,$$

donde $P(x)$ y $Q(x)$ son polinomios. No perdemos generalidad si suponemos que el numerador tiene grado menor que el denominador, pues en caso contrario podemos dividir:

$$P(x) = Q(x)C(x) + R(x),$$

donde $R(x)$ es nulo o tiene grado menor que $Q(x)$, y descomponer la integral como

$$\int \frac{P(x)}{Q(x)} \, dx = \int C(x) \, dx + \int \frac{R(x)}{Q(x)} \, dx.$$

También podemos suponer que el coeficiente director de $Q(x)$ vale 1 (si vale c , dividimos numerador y denominador entre c y así pasa a valer 1).

El teorema fundamental del álgebra [ITAn 3.27] afirma que todo polinomio con coeficientes complejos se descompone en la forma

$$Q(x) = (x - a_1) \cdots (x - a_n),$$

donde los a_i son números complejos (faltaría multiplicar por el coeficiente director, pero estamos suponiendo que vale 1).

Ahora bien, si $Q(x)$ tiene coeficientes reales, como sucede en el caso que nos ocupa y a_i es una raíz compleja, entonces, aplicando la conjugación a $Q(a_i) = 0$ obtenemos que $Q(\bar{a}_i) = 0$, luego el conjugado \bar{a}_i tiene que ser también raíz de Q . Y si agrupamos las raíces imaginarias en pares de conjugados, cada par a y \bar{a} determina un par de factores

$$(x - a)(x - \bar{a}) = x^2 - 2\operatorname{Re}(a)x - |a|^2$$

que dan lugar a un polinomio de grado 2 con coeficientes reales.

Concluimos entonces que todo polinomio $Q(x)$ (con coeficiente director 1) factoriza en la forma:

$$Q(x) = (x - a_1)^{n_1} \cdots (x - a_r)^{n_r} (x^2 + b_{r+1}x + c_{r+1})^{n_{t+1}} \cdots (x^2 + b_t x + c_t)^{n_t},$$

donde los factores son distintos dos a dos (simplemente porque hemos agrupado los que eran iguales).

Por ejemplo:

$$x^4 - 5x^3 + 4x^2 + 3x + 9 = (x - 3)^2 (x^2 + x + 1).$$

El problema de factorizar polinomios arbitrarios no es elemental en absoluto, pero es un problema algebraico del que no nos vamos a ocupar aquí. Si tenemos que integrar un cociente de polinomios, supondremos que hemos podido factorizar el denominador.

Fracciones simples El resultado fundamental es que todo cociente de la forma $P(x)/Q(x)$, donde el grado de P es menor que el de Q , se descompone en forma única como suma de fracciones simples, donde una *fracción simple* es un cociente de polinomios de una de las formas siguientes:

$$\frac{A}{(x - a_i)^k},$$

donde a_i es una de las raíces reales de $Q(x)$ y $0 \leq k \leq n_i$, o bien

$$\frac{Mx + N}{(x^2 + b_i x + c_i)^k},$$

donde $x^2 + b_i x + c_i$ es uno de los factores sin raíces reales en que se descompone el denominador $Q(x)$ e igualmente $k \leq n_i$.

No vamos a demostrar este resultado porque la prueba es laboriosa, pero hay que destacar que no necesitamos demostrarlo para aplicarlo al cálculo de primitivas, porque basta encontrar en cada caso particular la descomposición en factores simples y comprobar que es correcta, sin necesidad de justificar que es posible encontrarla en todos los casos.

Ejemplo Vamos a calcular

$$\int \frac{5x^3 - 22x^2 + 16x + 2}{(x - 3)^2 (x^2 + x + 1)} dx.$$

Para ello planteamos la descomposición del integrando en fracciones simples:

$$\frac{5x^3 - 22x^2 + 16x + 2}{(x - 3)^2 (x^2 + x + 1)} = \frac{A}{x - 3} + \frac{B}{(x - 3)^2} + \frac{Mx + N}{x^2 + x + 1}.$$

El teorema sobre descomposición en fracciones simples garantiza que tienen que existir números A , B , M , N que cumplan esta igualdad para todo x , pero no

necesitamos demostrar que tienen que existir, sino que nos basta con demostrar que existen encontrándolos. Para ello se suma el miembro derecho:

$$\frac{5x^3 - 22x^2 + 16x + 2}{(x-3)^2(x^2+x+1)} = \frac{A(x-3)(x^2+x+1) + B(x^2+x+1) + (Mx+N)(x-3)^2}{(x-3)^2(x^2+x+1)}$$

Y sólo tenemos que garantizar que los numeradores son iguales. Para ello basta dar valores a x y obtener así ecuaciones que deben cumplir los coeficientes. Por ejemplo, haciendo $x = 3$ queda

$$-13 = 13B,$$

de donde deducimos que $B = -1$. Si hacemos $x = 0, 1, -1$ queda

$$2 = -3A - 1 + 9N, \quad 1 = -6A - 3 + 4(M + N), \quad -41 = -4A - 1 + 16(N - M),$$

y al resolver el sistema de ecuaciones resulta $A = 2, M = 3, N = 1$, luego

$$\frac{5x^3 - 22x^2 + 16x + 2}{(x-3)^2(x^2+x+1)} = \frac{2}{x-3} - \frac{1}{(x-3)^2} + \frac{3x+1}{x^2+x+1}.$$

Notemos que los numeradores de ambos miembros son polinomios de grado 3 que coinciden en $x = 0, 1, -1, 3$, luego su diferencia tiene al menos cuatro raíces, luego tiene que ser el polinomio nulo. Así, hemos demostrado que la igualdad anterior se cumple para todo x (sin necesidad del teorema sobre fracciones simples). Por lo tanto,

$$\int \frac{5x^3 - 22x^2 + 16x + 2}{(x-3)^2(x^2+x+1)} dx = 2 \int \frac{1}{x-3} dx - \int \frac{1}{(x-3)^2} dx + \int \frac{3x+1}{x^2+x+1} dx.$$

Las primeras integrales son inmediatas:

$$2 \int \frac{1}{x-3} dx = 2 \log|x-3| + c, \quad - \int \frac{1}{(x-3)^2} dx = \frac{1}{x-3} + c,$$

y la última puede descomponerse en suma de una integral de tipo logaritmo más una de tipo arco tangente:

$$\int \frac{3x+1}{x^2+x+1} dx = \frac{3}{2} \int \frac{2x+1}{x^2+x+1} dx - \frac{1}{2} \int \frac{1}{x^2+x+1} dx.$$

La primera es

$$\frac{3}{2} \int \frac{2x+1}{x^2+x+1} dx = \frac{3}{2} \log(x^2+x+1) + c$$

y en la segunda completamos un cuadrado:

$$-\frac{1}{2} \int \frac{1}{x^2+x+1} dx = -\frac{1}{2} \int \frac{1}{(x+1/2)^2 + 3/4} dx =$$

$$\begin{aligned} -\frac{2}{3} \int \frac{1}{((2/\sqrt{3})x + 1/\sqrt{3})^2 + 1} dx &= -\frac{\sqrt{3}}{3} \int \frac{2/\sqrt{3}}{((2/\sqrt{3})x + 1/\sqrt{3})^2 + 1} dx \\ &= -\frac{\sqrt{3}}{3} \arctan \frac{1+2x}{\sqrt{3}} + c. \end{aligned}$$

En total:

$$\begin{aligned} \int \frac{5x^3 - 22x^2 + 16x + 2}{(x-3)^2(x^2+x+1)} dx &= 2 \log|x-3| + \frac{1}{x-3} + \frac{3}{2} \log(x^2+x+1) \\ &\quad - \frac{\sqrt{3}}{3} \arctan \frac{1+2x}{\sqrt{3}} + c. \end{aligned} \quad \blacksquare$$

En general, las fracciones simples correspondientes a raíces reales son inmediatas:

$$\int \frac{A}{x-a} dx = A \log|x-a| + c, \quad \int \frac{A}{(x-a)^{n+1}} dx = -\frac{A}{(x-a)^n} + c,$$

y las de la forma

$$\int \frac{Mx+N}{x^2+bx+c} dx,$$

como ilustra el ejemplo precedente, siempre pueden descomponerse como

$$\begin{aligned} \int \frac{Mx+N}{x^2+bx+c} dx &= \frac{M}{2} \int \frac{2x+b}{x^2+bx+c} dx - \int \frac{N-Mb/2}{x^2+bx+c} dx \\ &= \frac{M}{2} \log(x^2+bx+c) + \frac{Mb-2N}{2} \int \frac{1}{(x-b/2)^2 + (4c-b^2)/4} dx \\ &= \frac{M}{2} \log(x^2+bx+c) + \frac{Mb-2N}{2} \int \frac{1}{\left(\frac{2x-b}{\sqrt{4c-b^2}}\right)^2 + 1} dx \\ &= \frac{M}{2} \log(x^2+bx+c) + \frac{Mb-2N}{2} \int \frac{1}{\left(\frac{2x-b}{\sqrt{4c-b^2}}\right)^2 + 1} dx \\ &= \frac{M}{2} \log(x^2+bx+c) + \frac{(Mb-2N)\sqrt{4c-b^2}}{4} \int \frac{2/\sqrt{4c-b^2}}{\left(\frac{2x-b}{\sqrt{4c-b^2}}\right)^2 + 1} dx \\ &= \frac{M}{2} \log(x^2+bx+c) + \frac{(Mb-2N)\sqrt{4c-b^2}}{4} \arctan \frac{2x-b}{\sqrt{4c-b^2}} + c. \end{aligned}$$

Nos falta considerar las de la forma:

$$\begin{aligned} \int \frac{Mx+N}{(x^2+bx+c)^{n+1}} dx &= \frac{M}{2} \int \frac{2x+b}{(x^2+bx+c)^{n+1}} dx - \int \frac{N-Mb/2}{(x^2+bx+c)^{n+1}} dx \\ &= -\frac{M}{2} \frac{1}{(x^2+bx+c)^n} + \frac{Mb-2N}{2} \int \frac{1}{(x^2+bx+c)^{n+1}} dx, \end{aligned}$$

luego sólo queda pendiente cómo integrar las fracciones racionales de la forma

$$\int \frac{1}{(x^2 + bx + c)^{n+1}} dx,$$

donde el denominador no tiene raíces reales. Más adelante veremos una forma de tratarlas (C.12),(C.17), pero por completitud vamos a describir un procedimiento que, más que resolverlas, nos evita llegar a ellas.

El método de Hermite El método de Hermite es una variante del método de descomposición en fracciones simples, que es bastante más laborioso, pero que evita la aparición de fracciones simples con raíces múltiples.

Se basa en el hecho de que una fracción $P(x)/Q(x)$ con el grado de P menor que el de Q y donde

$$Q(x) = (x - a_1)^{n_1} \cdots (x - a_r)^{n_r} (x^2 + b_{r+1}x + c_{r+1})^{n_{t+1}} \cdots (x^2 + b_t x + c_t)^{n_t},$$

puede descomponerse en suma de fracciones simples de la forma

$$\frac{A}{x - a_i}, \quad \frac{Mx + n}{x^2 + b_i x + c_i}$$

más un último término de la forma

$$\frac{d}{dx} \frac{P_1(x)}{Q_1(x)},$$

donde

$$Q_1(x) = (x - a_1)^{n_1 - 1} \cdots (x - a_r)^{n_r - 1} (x^2 + b_{r+1}x + c_{r+1})^{n_{t+1} - 1} \cdots (x^2 + b_t x + c_t)^{n_t - 1},$$

(de modo que las raíces con multiplicidad 1 desaparecen) y $P_1(x)$ tiene grado menor que $Q_1(x)$.

Nuevamente, no vamos a demostrar que siempre existe una descomposición en estas condiciones, pero no hace falta demostrarlo para resolver sin lagunas lógicas cualquier integral concreta.

Ejemplo Para calcular la integral

$$\int \frac{x^2 - 2}{x^3(x^2 + 1)^2} dx$$

planteamos la descomposición

$$\frac{x^2 - 2}{x^3(x^2 + 1)^2} = \frac{A}{x} + \frac{Mx + N}{x^2 + 1} + \frac{d}{dx} \frac{B_3x^3 + B_2x^2 + B_1x + B_0}{x^2(x^2 + 1)}.$$

Para calcular los coeficientes tenemos que calcular la derivada:

$$\frac{(3B_3x^2 + 2B_2x + B_1)x(x^2 + 1) - (B_3x^3 + B_2x^2 + B_1x + B_0)(4x^2 + 2)}{x^3(x^2 + 1)^2},$$

sumar las tres fracciones e igualar los numeradores:

$$\begin{aligned} x^2 - 2 &= Ax^2(x^2 + 1)^2 + (Mx + N)x^3(x^2 + 1) \\ &+ (3B_3x^2 + 2B_2x + B_1)x(x^2 + 1) - (B_3x^3 + B_2x^2 + B_1x + B_0)(4x^2 + 2). \end{aligned}$$

En principio, dando siete valores a x obtenemos un sistema de siete ecuaciones con siete incógnitas que nos determinará los coeficientes, aunque es posible obtener algunos más rápidamente.

Por ejemplo, haciendo $x = 0$ obtenemos $-2 = -2B_0$, luego $B_0 = 1$. Igualando los coeficientes de grado 1 queda

$$0 = 0 + 0 + B_1 - 2B_1 = -B_1,$$

luego $B_1 = 0$. Igualando los coeficientes de grado 2 queda

$$1 = A + 2B_2 - 4B_0 - 2B_2 = A - 4,$$

luego $A = 5$. Igualando los coeficientes de grado 6 obtenemos $A + M = 0$, luego $M = -A = -5$. Hasta aquí tenemos:

$$\begin{aligned} x^2 - 2 &= 5(x^6 + 2x^4 + x^2) + (-5x + N)(x^5 + x^3) \\ &+ (3B_3x + 2B_2)(x^4 + x^2) - (B_3x^3 + B_2x^2 + 1)(4x^2 + 2). \end{aligned}$$

Igualando los coeficientes de grado 4 queda $0 = 10 - 5 + 2B_2 - 4B_2$, luego $B_2 = 5/2$. Si ahora hacemos $x = 1$ queda

$$-1 = 20 + (2N - 10) + (6B_3 + 10) - (6B_3 + 15 + 6) = 2N - 1,$$

luego $N = 0$, con lo que tenemos

$$x^2 - 2 = 5x^4 + 5x^2 + (3B_3x + 5)(x^4 + x^2) - (B_3x^3 + (5/2)x^2 + 1)(4x^2 + 2),$$

y el coeficiente de x^3 es $B_3 = 0$. En conclusión:

$$\frac{x^2 - 2}{x^3(x^2 + 1)^2} = \frac{5}{x} - \frac{5x}{x^2 + 1} + \frac{d}{dx} \frac{(5/2)x^2 + 1}{x^2(x^2 + 1)},$$

de donde, a su vez,

$$\begin{aligned} \int \frac{x^2 - 2}{x^3(x^2 + 1)^2} dx &= \frac{5x^2 + 2}{2x^2(x^2 + 1)} + 5 \int \frac{1}{x} dx - 5 \int \frac{x}{x^2 + 1} dx \\ &= \frac{5x^2 + 2}{2x^2(x^2 + 1)} + 5 \log |x| - \frac{5}{2} \log(x^2 + 1) + c. \end{aligned}$$

■

C.3 Integración por partes

Veamos ejemplos de cálculo de primitivas mediante la fórmula de integración por partes:¹

$$\int u dv = uv - \int v du.$$

(Véase la página 2.2 para su justificación.) Un ejemplo típico de aplicación es el primero de la lista siguiente:

Ejemplos Veamos que

$$\int x^2 \operatorname{sen} x dx = -x^2 \cos x + 2x \operatorname{sen} x + 2 \cos x + c. \quad (\text{C.3})$$

La fórmula de integración por partes nos permite sustituir uno de los factores por su derivada y el otro por su integral. Al derivar o integrar el seno obtenemos un coseno que es “esencialmente lo mismo”, pero al derivar x^2 pasamos a $2x$ y, si lo volvemos a derivar, pasamos a 2, lo que sugiere que aplicando dos veces la integración por partes llegaremos a una integral inmediata. En efecto:

$$u = x^2, \quad dv = \operatorname{sen} x dx, \quad du = 2x dx, \quad v = \int \operatorname{sen} x dx = -\cos x.$$

$$\int x^2 \operatorname{sen} x dx = -x^2 \cos x + \int 2x \cos x dx$$

y de nuevo:

$$u = 2x, \quad dv = \cos x dx, \quad du = 2 dx, \quad v = \int \cos x dx = \operatorname{sen} x.$$

$$= -x^2 \cos x + 2x \operatorname{sen} x - \int 2 \operatorname{sen} x dx = -x^2 \cos x + 2x \operatorname{sen} x + 2 \cos x + c.$$

■

La idea general es derivar una parte de la integral (la que llamamos u) para simplificarla. Por ejemplo, en la integral siguiente llamamos $u = \log x$ porque así, al derivarlo, en la integral queda $1/x$, que es más sencillo:

$$\int x \log x dx.$$

Aplicamos la fórmula de integración por partes con $u = \log x$ y $dv = x dx$, con lo que

$$du = \frac{1}{x} dx, \quad v = \int x dx = \frac{x^2}{2}$$

y así

$$\begin{aligned} \int x \log x dx &= \frac{x^2}{2} \log x - \int \frac{x^2}{2} \frac{1}{x} dx = \frac{x^2}{2} \log x - \frac{1}{2} \int x dx \\ &= \frac{x^2}{2} \log x - \frac{x^2}{4} + c. \end{aligned}$$

■

¹Es habitual recordarla con la frase “Un día vi una vaca vestida de uniforme.”

Nada impide aplicar la regla tomando como $dv = 1 \cdot dx$, con lo que $v = x$, que es lo que conviene para calcular:

$$\int \log x \, dx = x \log x - \int x \frac{1}{x} \, dx = x \log x - \int dx = x \log x - x + c. \quad (\text{C.4})$$

■

La misma técnica puede aplicarse para calcular

$$\int \arctan x \, dx,$$

llamando

$$u = \arctan x, \quad dv = dx, \quad du = \frac{1}{1+x^2} \, dx, \quad v = \int dx = x.$$

Al aplicar la fórmula de integración por partes queda

$$\begin{aligned} \int \arctan x \, dx &= x \arctan x - \int \frac{x}{1+x^2} \, dx = x \arctan x - \frac{1}{2} \int \frac{2x}{1+x^2} \, dx \\ &= x \arctan x - \frac{1}{2} \log(1+x^2) + c. \end{aligned}$$

■

A veces, al aplicar una o más veces la fórmula de integración por partes obtenemos la misma integral de partida, en cuyo caso podemos calcularla despejándola. Veamos algunos ejemplos:

$$\begin{aligned} \int \sen t e^{-st} \, dt &= -\frac{e^{-st}}{1+s^2} (\cos t + s \sen t) + c, \\ \int \cos t e^{-st} \, dt &= \frac{e^{-st}}{1+s^2} (\sen t - s \cos t) + c. \end{aligned} \quad (\text{C.5})$$

Para calcular la primera integral tomamos

$$u = e^{-st}, \quad du = -se^{-st} \, dt, \quad dv = \sen t \, dt, \quad v = \int \cos t \, dt = -\cos t.$$

Entonces

$$\int \sen t e^{-st} \, dt = -\cos t e^{-st} - s \int \cos t e^{-st} \, dt.$$

La integral resultante la calculamos también por partes:

$$u = e^{-st}, \quad du = -se^{-st} \, dt, \quad dv = \cos t \, dt, \quad v = \int \cos t \, dt = \sen t.$$

Así:

$$\int \sen t e^{-st} \, dt = -\cos t e^{-st} - s \sen t e^{-st} - s^2 \int \sen t e^{-st} \, dt.$$

Despejando:

$$(1 + s^2) \int \operatorname{sen} t e^{-st} dt = -e^{-st}(\cos t + s \operatorname{sen} t),$$

de donde se sigue la primera igualdad, y la segunda se razona análogamente. ■

$$\int x \operatorname{sen}^2 x dx = \frac{x^2}{4} - \frac{x}{4} \operatorname{sen} 2x - \frac{1}{8} \cos 2x + c. \quad (\text{C.6})$$

Para calcular esta integral podríamos llamar $dv = \operatorname{sen} x^2 dx$ y usar (C.2) para calcular v , pero una alternativa más sencilla es hacer:

$$u = x \operatorname{sen} x, \quad dv = \operatorname{sen} x dx, \quad du = (\operatorname{sen} x + x \cos x) dx, \quad v = -\cos x.$$

$$\begin{aligned} \int x \operatorname{sen}^2 x dx &= -x \operatorname{sen} x \cos x + \int (\cos x \operatorname{sen} x + x \cos^2 x) dx \\ &= -\frac{x}{2} \operatorname{sen} 2x + \frac{1}{2} \int \operatorname{sen} 2x dx + \int x(1 - \operatorname{sen}^2 x) dx \\ &= -\frac{x}{2} \operatorname{sen} 2x - \frac{1}{4} \cos 2x + \frac{x^2}{2} - \int x \operatorname{sen}^2 x. \end{aligned}$$

Como hemos obtenido la misma integral inicial, podemos despejarla y obtenemos el resultado anunciado. ■

Similarmente podemos calcular

$$\int \sqrt{r^2 - x^2} dx = \frac{x}{2} \sqrt{r^2 - x^2} - \frac{r^2}{2} \arccos \frac{x}{r} + c. \quad (\text{C.7})$$

Tomamos

$$u = \sqrt{r^2 - x^2}, \quad dv = dx, \quad du = -\frac{x}{\sqrt{r^2 - x^2}} dx, \quad v = x.$$

Así:

$$\begin{aligned} \int \sqrt{r^2 - x^2} dx &= x \sqrt{r^2 - x^2} + \int \frac{x^2}{\sqrt{r^2 - x^2}} dx \\ &= x \sqrt{r^2 - x^2} + \int \frac{r^2}{\sqrt{r^2 - x^2}} dx - \int \frac{r^2 - x^2}{\sqrt{r^2 - x^2}} dx \\ &= x \sqrt{r^2 - x^2} - r^2 \int \frac{-1/r}{\sqrt{1 - (x/r)^2}} dx - \int \sqrt{r^2 - x^2} dx, \end{aligned}$$

y al despejar se obtiene la función indicada. Del mismo modo podemos calcular:

$$\int \sqrt{1 + x^2} dx = \frac{x}{2} \sqrt{1 + x^2} + \frac{1}{2} \operatorname{arsenh} x + c. \quad (\text{C.8})$$

■

Otro ejemplo notable de integración cíclica es el siguiente:

$$\int \operatorname{sen}^n x \, dx = -\frac{1}{n} \operatorname{sen}^{n-1} x \cos x + \frac{n-1}{n} \int \operatorname{sen}^{n-2} x \, dx, \quad (\text{C.9})$$

$$\int \cos^n x \, dx = \frac{1}{n} \cos^{n-1} x \operatorname{sen} x + \frac{n-1}{n} \int \cos^{n-2} x \, dx.$$

Separamos

$$\int \operatorname{sen}^n x \, dx = \int \operatorname{sen}^{n-1} x \operatorname{sen} x \, dx$$

y aplicamos la fórmula de integración por partes con

$$u = \operatorname{sen}^{n-1} x, \quad dv = \operatorname{sen} x \, dx, \quad du = (n-1) \operatorname{sen}^{n-2} x \cos x \, dx, \quad v = -\cos x.$$

Así (cambiando $\cos^2 x = 1 - \operatorname{sen}^2 x$):

$$\begin{aligned} \int \operatorname{sen}^n x \, dx &= -\operatorname{sen}^{n-1} x \cos x + (n-1) \int \operatorname{sen}^{n-2} x \cos^2 x \, dx \\ &= -\operatorname{sen}^{n-1} x \cos x + (n-1) \int \operatorname{sen}^{n-2} x \, dx - (n-1) \int \operatorname{sen}^n x \, dx, \end{aligned}$$

y al despejar se obtiene la fórmula indicada. La fórmula para el coseno se prueba análogamente. ■

Funciones racionales Veamos cómo la integración por partes nos permite calcular integrales de funciones racionales con raíces complejas múltiples sin necesidad de aplicar el método de Hermite. Consideremos en primer lugar el caso siguiente:

$$\int \frac{x^2}{(ax^2 + c)^{n+1}} \, dx.$$

No es evidente que el método de partes vaya a ser de utilidad, pero lo cierto es que podemos aplicarlo con

$$u = x, \quad dv = \frac{x}{(ax^2 + c)^{n+1}} \, dx, \quad v = \int \frac{x}{(ax^2 + c)^{n+1}} \, dx = -\frac{1}{2an} \frac{1}{(ax^2 + c)^n}.$$

Así:

$$\int \frac{x^2}{(ax^2 + c)^{n+1}} \, dx = -\frac{1}{2an} \frac{x}{(ax^2 + c)^n} + \frac{1}{2an} \int \frac{1}{(ax^2 + c)^n} \, dx. \quad (\text{C.10})$$

Si $n = 1$ y $c > 0$, la integral resultante es inmediata de tipo arco tangente y tenemos resuelta la integral. En particular:

$$\int \frac{x^2}{(x^2 + 1)^2} \, dx = -\frac{x}{2(x^2 + 1)} + \frac{1}{2} \arctan x + c. \quad (\text{C.11})$$

Pero podemos usar el caso general para integrar una fracción simple de la forma

$$\begin{aligned} \int \frac{1}{(ax^2 + c)^{n+1}} dx &= \frac{1}{c} \int \frac{ax^2 + c}{(ax^2 + c)^{n+1}} dx - \frac{a}{c} \int \frac{x^2}{(ax^2 + c)^{n+1}} dx \\ &= \frac{1}{c} \int \frac{1}{(ax^2 + c)^n} dx + \frac{1}{2cn} \frac{x}{(ax^2 + c)^n} - \frac{1}{2cn} \int \frac{1}{(ax^2 + c)^n} dx, \end{aligned}$$

luego

$$\int \frac{1}{(ax^2 + c)^{n+1}} dx = \frac{1}{2cn} \frac{x}{(ax^2 + c)^n} + \frac{2n-1}{2cn} \int \frac{1}{(ax^2 + c)^n} dx, \quad (\text{C.12})$$

y esta fórmula nos permite ir reduciendo el grado del denominador hasta que la integral sea de tipo arco tangente.

Necesitaremos este caso particular, que se descompone en integrales de los dos tipos que hemos considerado, (C.10) y (C.12):

$$\begin{aligned} &\int \frac{1+x^2}{(1+\epsilon+(1-\epsilon)x^2)^2} dx \\ &= \int \frac{1}{(1+\epsilon+(1-\epsilon)x^2)^2} dx + \int \frac{x^2}{(1+\epsilon+(1-\epsilon)x^2)^2} dx \\ &= \frac{1}{2(1+\epsilon)} \frac{x}{1+\epsilon+(1-\epsilon)x^2} + \frac{1}{2(1+\epsilon)} \int \frac{1}{1+\epsilon+(1-\epsilon)x^2} dx \\ &\quad - \frac{1}{2(1-\epsilon)} \frac{x}{1+\epsilon+(1-\epsilon)x^2} + \frac{1}{2(1-\epsilon)} \int \frac{1}{1+\epsilon+(1-\epsilon)x^2} dx \\ &= -\frac{\epsilon}{1-\epsilon^2} \frac{x}{1+\epsilon+(1-\epsilon)x^2} + \frac{1}{1-\epsilon^2} \int \frac{1}{1+\epsilon+(1-\epsilon)x^2} dx, \end{aligned}$$

y la última integral puede reducirse a un arco tangente:

$$\begin{aligned} &\int \frac{1}{1+\epsilon+(1-\epsilon)x^2} dx = \frac{1}{1+\epsilon} \int \frac{1}{1+\left(\sqrt{\frac{1-\epsilon}{1+\epsilon}}x\right)^2} dx \\ &= \frac{1}{1+\epsilon} \sqrt{\frac{1+\epsilon}{1-\epsilon}} \int \frac{\sqrt{\frac{1-\epsilon}{1+\epsilon}}}{1+\left(\sqrt{\frac{1-\epsilon}{1+\epsilon}}x\right)^2} ds = \frac{1}{\sqrt{1-\epsilon^2}} \arctan\left(\sqrt{\frac{1-\epsilon}{1+\epsilon}}x\right) + c. \end{aligned}$$

En total:

$$\begin{aligned} \int \frac{1+x^2}{(1+\epsilon+(1-\epsilon)x^2)^2} dx &= -\frac{\epsilon}{1-\epsilon^2} \frac{x}{1+\epsilon+(1-\epsilon)x^2} \\ &\quad + \frac{1}{1-\epsilon^2} \frac{1}{\sqrt{1-\epsilon^2}} \arctan\left(\sqrt{\frac{1-\epsilon}{1+\epsilon}}x\right) + c. \end{aligned} \quad (\text{C.13})$$

C.4 Integración por sustitución

Veamos ahora ejemplos de integración por cambio de variable, tal y como se explica en la página 2.2.

Ejemplo Vamos a calcular de nuevo la integral (C.7):

$$\int \sqrt{r^2 - x^2} dx = \frac{x}{2} \sqrt{r^2 - x^2} - \frac{r^2}{2} \arccos \frac{x}{r} + c. \quad (\text{C.14})$$

Notemos que el integrando está definido en $]-r, r[$, en el cual es biyectiva la función $x(t) = r \cos t$, luego podemos usarla para hacer un cambio de variable, con $dx = -r \sin t dt$. Por lo tanto:

$$\begin{aligned} \int \sqrt{r^2 - x^2} dx &= -r \int \sqrt{r^2(1 - \cos^2 t)} \sin t dt = -r^2 \int \sin^2 t dt \\ &= -r^2 \left(\frac{t}{2} - \frac{1}{4} \sin 2t + c \right) = -\frac{r^2}{2} (t - \sin t \cos t + c) \\ &= -\frac{r^2}{2} (t - \cos t \sqrt{1 - \cos^2 t} + c) \end{aligned}$$

Ahora falta deshacer el cambio de variable, $\cos t = x/r$:

$$\begin{aligned} \int \sqrt{r^2 - x^2} dx &= -\frac{r^2}{2} \left(\arccos \frac{x}{r} - \frac{x}{r} \sqrt{1 - x^2/r^2} \right) + c \\ &= -\frac{r^2}{2} \arccos \frac{x}{r} + \frac{x}{2} \sqrt{r^2 - x^2} + c, \end{aligned}$$

con lo que obtenemos de nuevo el resultado de (C.7). ■

Ejemplo Vamos a comprobar que, si $a > 0$,

$$\int \sqrt{\frac{x}{a-x}} dx = a \arcsen \sqrt{\frac{x}{a}} - \sqrt{x(a-x)} + c. \quad (\text{C.15})$$

Notemos que el integrando está definido en el intervalo $[0, a]$, y podemos considerar la función $x : [0, \pi/2] \rightarrow [0, a]$ dada por $x(t) = a \sin^2 t$, que tiene inversa $t(x) = \arcsen \sqrt{x/a}$. Sustituimos $dx = 2a \sin t \cos t dt$, con lo que queda:

$$\begin{aligned} \int \sqrt{\frac{a \sin^2 t}{a - a \sin^2 t}} 2a \sin t \cos t dt &= \int \frac{\sin t}{\cos t} 2a \sin t \cos t dt = 2a \int \sin^2 t dt \\ &= at - \frac{a}{2} \sin 2t + c = at - a \sin t \cos t + c, \end{aligned}$$

donde hemos usado (C.2). Para deshacer el cambio de variable observamos que

$$\cos t = \sqrt{1 - \sin^2 t} = \sqrt{1 - x/a} = \sqrt{\frac{a-x}{a}},$$

luego

$$\begin{aligned}\int \sqrt{\frac{x}{a-x}} dx &= a \operatorname{arcsen} \sqrt{\frac{x}{a}} - a \sqrt{\frac{x}{a}} \sqrt{\frac{a-x}{a}} + c \\ &= a \operatorname{arcsen} \sqrt{\frac{x}{a}} - \sqrt{x(a-x)} + c. \quad \blacksquare\end{aligned}$$

La sustitución de Weierstrass Cuando el integrando está formado únicamente por senos y cosenos operados algebraicamente, es decir, cuando es de la forma $R(\operatorname{sen} x, \operatorname{cos} x)$, donde $R(u, v)$ es un cociente de polinomios, hay un cambio de variable que transforma la integral en otra cuyo integrando es un cociente de polinomios:

$$u = \tan \frac{x}{2}, \quad \operatorname{cos} x = \frac{1-u^2}{1+u^2}, \quad \operatorname{sen} x = \frac{2u}{1+u^2}, \quad dx = \frac{2}{1+u^2} du.$$

En efecto, la fórmula para dx se obtiene derivando $x = 2 \operatorname{arctan} u$ y las otras dos son trigonometría:

$$\operatorname{cos} x = \operatorname{cos}^2 \frac{x}{2} - \operatorname{sen}^2 \frac{x}{2} = 2 \operatorname{cos}^2 \frac{x}{2} - 1 = 2 \frac{1}{1+u^2} - 1 = \frac{1-u^2}{1+u^2},$$

donde hemos usado que al dividir $\operatorname{sen}^2(x/2) + \operatorname{cos}^2(x/2) = 1$ entre $\operatorname{cos}^2(x/2)$ nos da $u^2 + 1 = 1/\operatorname{cos}^2(x/2)$.

$$\operatorname{sen} x = 2 \operatorname{sen} \frac{x}{2} \operatorname{cos} \frac{x}{2} = 2 \tan \frac{x}{2} \operatorname{cos}^2 \frac{x}{2} = 2u \frac{1}{1+u^2}.$$

En muchos casos existen cambios de variable alternativos que dan lugar a integrales más sencillas. Concretamente:

- Si el integrando es impar en el seno, es decir, si se cumple

$$R(-\operatorname{sen} x, \operatorname{cos} x) = -R(\operatorname{sen} x, \operatorname{cos} x),$$

sirve el cambio:

$$\operatorname{cos} x = u, \quad \operatorname{sen} x = \sqrt{1-u^2}, \quad dx = -\frac{1}{\sqrt{1-u^2}} du.$$

- Si el integrando es impar en el coseno ($R(\operatorname{sen} x, -\operatorname{cos} x) = -R(\operatorname{sen} x, \operatorname{cos} x)$) sirve el cambio:

$$\operatorname{sen} x = u, \quad \operatorname{cos} x = \sqrt{1-u^2}, \quad dx = \frac{1}{\sqrt{1-u^2}} du.$$

- Si $R(-\operatorname{sen} x, -\operatorname{cos} x) = R(\operatorname{sen} x, \operatorname{cos} x)$, sirve el cambio

$$\tan x = u, \quad \operatorname{cos} x = \frac{1}{\sqrt{1+u^2}}, \quad \operatorname{sen} x = \frac{u}{\sqrt{1+u^2}}, \quad dx = \frac{1}{1+u^2} du.$$

En todos los casos la hipótesis correspondiente hace que las raíces se simplifiquen al hacer la sustitución y desaparezcan. Podríamos demostrarlo en general, pero basta comprobar que se cumple en cada caso particular.

Ejemplo Para calcular la integral

$$\int \frac{1}{\operatorname{sen} x} dx = \log \sqrt{\frac{1 - \cos x}{1 + \cos x}} + c$$

aplicamos el cambio primero de los tres cambios alternativos, con lo que

$$\begin{aligned} \int \frac{1}{\operatorname{sen} x} dx &= - \int \frac{1}{\sqrt{1-u^2}} \frac{1}{\sqrt{1-u^2}} du = - \int \frac{1}{1-u^2} du \\ &= \frac{1}{2} \int \frac{1}{u-1} du - \frac{1}{2} \int \frac{1}{u+1} du = \frac{1}{2} \log |u-1| - \frac{1}{2} \log |u+1| + c \\ &= \log \sqrt{\frac{|1-u|}{|1+u|}} + c, \end{aligned}$$

y al deshacer el cambio de variable obtenemos el resultado indicado. ■

Ejemplo En el estudio de las leyes de Kepler en la sección 2.3 nos aparecerá la integral:

$$\frac{1}{2} \int \frac{1}{(1 + \epsilon \cos \theta)^2} d\theta, \quad 0 \leq \epsilon < 1.$$

Si aplicamos el cambio de variable $x = \tan(\theta/2)$:

$$\int \frac{1}{(1 + \epsilon \frac{1-x^2}{1+x^2})^2} \frac{1}{1+x^2} dx = \int \frac{1+x^2}{((1+\epsilon) + (1-\epsilon)x^2)^2} dx,$$

obtenemos precisamente (C.13), luego el resultado es

$$-\frac{\epsilon}{1-\epsilon^2} \frac{x}{1+\epsilon+(1-\epsilon)x^2} + \frac{1}{1-\epsilon^2} \frac{1}{\sqrt{1-\epsilon^2}} \arctan \left(\sqrt{\frac{1-\epsilon}{1+\epsilon}} x \right) + c.$$

Ahora deshacemos el cambio de variable:

$$\begin{aligned} \frac{1}{2} \int \frac{1}{(1 + \epsilon \cos \theta)^2} d\theta &= -\frac{\epsilon}{1-\epsilon^2} \frac{\tan(\theta/2)}{1+\epsilon+(1-\epsilon)\tan^2(\theta/2)} \\ &\quad + \frac{1}{(1-\epsilon^2)\sqrt{1-\epsilon^2}} \arctan \left(\sqrt{\frac{1-\epsilon}{1+\epsilon}} \tan \frac{\theta}{2} \right) + c. \end{aligned}$$

En el primer término podemos multiplicar numerador y denominador por $\cos^2(\theta/2)$ y queda

$$-\frac{\epsilon}{1-\epsilon^2} \frac{\operatorname{sen}(\theta/2) \cos(\theta/2)}{(1+\epsilon) \cos^2(\theta/2) + (1-\epsilon) \operatorname{sen}^2(\theta/2)} = -\frac{1}{2(1-\epsilon^2)} \frac{\epsilon \operatorname{sen} \theta}{1+\epsilon \cos \theta}.$$

Así pues:

$$\frac{1}{2} \int \frac{1}{(1 + \epsilon \cos \theta)^2} d\theta = \tag{C.16}$$

$$\frac{1}{(1 - \epsilon^2)\sqrt{1 - \epsilon^2}} \arctan \left(\sqrt{\frac{1 - \epsilon}{1 + \epsilon}} \frac{\operatorname{sen} \theta}{1 + \cos \theta} \right) - \frac{1}{2(1 - \epsilon^2)} \frac{\epsilon \operatorname{sen} \theta}{1 + \epsilon \cos \theta} + c.$$

■

Las sustituciones de Weierstrass también funcionan con senos y cosenos hiperbólicos:

Ejemplo

$$\int \frac{\operatorname{senh}^2 x}{\cosh^3 x} dx = -\frac{\tanh x}{2 \cosh x} + \frac{1}{2} \arctan \operatorname{senh} x + c.$$

Como es impar en el coseno hiperbólico, sustituimos

$$\operatorname{senh} x = t, \quad \cosh x = \sqrt{1 + t^2}, \quad \cosh x dx = dt.$$

$$\int \frac{\operatorname{senh}^2 x}{\cosh^3 x} dx = \int \frac{t^2}{(1 + t^2)^2} dt = -\frac{t}{2(1 + t^2)} + \frac{1}{2} \arctan t + c,$$

donde hemos usado (C.11). Al deshacer el cambio queda

$$\int \frac{\operatorname{senh}^2 x}{\cosh^3 x} dx = -\frac{\operatorname{senh} x}{2(1 + \operatorname{senh}^2 x)} + \frac{1}{2} \arctan \operatorname{senh} x + c.$$

■

Fracciones simples En (C.12) hemos visto cómo integrar fracciones simples de la forma

$$\frac{1}{(x^2 + c)^{n+1}}.$$

Un simple cambio de variable reduce a este caso el caso general:

$$\int \frac{1}{(x^2 + bx + c)^{n+1}} dx = \int \frac{1}{((x + b/2)^2 + c - b^2/4)^{n+1}} dx$$

$$= \int \frac{1}{(y^2 + c - b^2/4)^{n+1}} dy, \tag{C.17}$$

donde $y = x + b/2$, con lo cual tenemos un método para integrar cualquier función racional por descomposición en fracciones simples sin necesidad de aplicar el método de Hermite.

■

Integrales binomias Una integral binomia es una integral de la forma

$$\int x^m (a + bx^n)^p dx,$$

donde a, b son números reales y m, n, p son números racionales. En los casos siguientes pueden reducirse a integrales de funciones racionales mediante los cambios de variable indicados:

- Si p es entero, hacemos $x = t^s$, donde s es el mínimo común múltiplo de los denominadores de m y n .
- Si $(m+1)/n$ es entero, hacemos $a + bx^n = t^s$, donde s es el denominador de p .
- Si $p + (m+1)/n$ es entero, hacemos $ax^{-n} + b = t^s$, donde s es el denominador de p .

Es fácil justificar que los cambios indicados llevan siempre a una función racional, aunque no es necesario justificarlo en general para aplicar el método en cada caso en particular. No obstante, veamos como ilustración el tercer caso:

Si $ax^{-n} + b = t^s$, entonces $x^n = a(t^s - b)^{-1}$, luego $x = a^{1/n}(t^s - b)^{-1/n}$, luego

$$dx = -(1/n)a^{1/n}(t^s - b)^{-1/n-1}st^{s-1} dt$$

y $a + bx^n = (ax^{-n} + b)x^n = t^s a(t^s - b)^{-1}$, luego la integral se convierte en

$$-\frac{sa^{m/n+p+1/n}}{n} \int (t^s - b)^{-p-(m+1)/n-1} t^{ps+s-1} dt$$

y todos los exponentes en el integrando son enteros, luego se trata de un cociente de polinomios. ■

Ejemplo La integral

$$\int (1 + x^2)^{-3/2} dx$$

es binomia con $m = 0$, $n = 2$, $p = -3/2$ y está en el tercer caso, pues se cumple que $p + (m+1)/n = -1$. Por consiguiente, aplicamos el cambio de variable

$$x^{-2} + 1 = t^2.$$

Así, $x = (t^2 - 1)^{-1/2}$, luego

$$dx = -(t^2 - 1)^{-3/2} t dt, \quad 1 + x^2 = x^2(x^{-2} + 1) = (t^2 - 1)^{-1} t^2,$$

luego

$$\int (1 + x^2)^{-3/2} dx = - \int (t^2 - 1)^{3/2} t^{-3} (t^2 - 1)^{-3/2} t dt = - \int \frac{1}{t^2} dt = \frac{1}{t} + c$$

y, deshaciendo el cambio de variable,

$$\int \frac{1}{\sqrt{(1+x^2)^3}} dx = \frac{1}{\sqrt{x^{-2}+1}} + c = \frac{x}{\sqrt{1+x^2}} + c. \quad (\text{C.18})$$

■

Índice de Materias

- abierto, 3
- aceleración, 202
 - angular, 220
 - sobre la trayectoria, 202
 - tangencial, normal, 207
- amplitud de Jacobi, 376
- arco, 98
 - rectificable, 99
 - seno/coseno lemniscáticos, 369

- Barrow (regla de), 425
- Beaume (problema de), 126
- binormal (vector), 210

- cóncava (función), 31
- cambio de parámetro, 102
- campo vectorial, 329
 - conservativo, 330
- cardioide, 472
- catenaria, 229
- Cauchy (fórmulas integrales), 437
- Cauchy-Riemann (ecuaciones de), 419
- Cavalieri (principio de), 117
- celda, 140
- centrífuga (fuerza), 265
- cicloide, 467
- circunferencia osculatriz, 210
- clotoide, 415
- conexo (abierto), 445
- convexa (función), 31
- convexo (conjunto), 433
- convolución, 187
- Coriolis (fuerza de), 265
- criterio
 - de Cauchy, 151
 - de Dirichlet, 153
- curvatura, 204

- derivable
 - curva paramétrica, 9
 - función, 3, 5
- derivada, 3, 9
 - direccional, 285
- determinante, 464
- diferenciable (función), 282
- diferencial, 65, 285
- Dirichlet (integral de), 179

- ecuación diferencial, 52
- energía
 - cinética, 326
 - mecánica, 330
 - potencial, 330
- entera (función), 416
- escalar, 456
- espiral
 - de Arquímedes, 106
 - de Euler, 415
- estrellado (conjunto), 433
- Euler
 - espiral de, 415
 - fuerza de, 265
 - integral de, 196
- exponencialmente acotada (función), 172

- factorial (función), 188
- figura elemental, 140
- fracciones simples, 493
- Frenet (fórmulas de), 211
- Fresnel (integrales de), 154

- Gabriel (cuerno de), 122
- gradiente, 314
- gravitación universal, 236

- hélice, 9
- holomorfa (función), 416
- Hooke (ley de), 232
- integral
 - curvilínea, 422
 - definida, 70
 - elíptica
 - de primera especie, 356
 - de segunda especie, 354
 - impropia, 144
 - indefinida, 69
- intensidad del campo gravitatorio, 223
- isócrona (curva), 332
- Jacobi (funciones elípticas de), 376
- jacobiana (matriz), 419
- Kepler (leyes de)
 - primera, 92
 - segunda, 93
 - tercera, 95
- Laplace (transformada de), 171
- Legendre
 - fórmula de duplicación, 197
 - relación de, 362
- Leibniz (regla de), 305
- lemniscata, 481
- logística
 - ecuación diferencial, 127
 - función, 128
- longitud de un arco, 102
- máximo local, 22, 289
- media aritmético-geométrica, 403
- mínimo local, 22, 289
- nefroide, 474
- Newton (leyes de)
 - primera, 221
 - segunda, 222
 - tercera, 224
- norma, 455
- normal (vector), 458
- normal unitario (vector), 207
- ortogonales (vectores), 457
- oscilador armónico, 58
- óvalo de Cassini, 483
- paralelas, 458
- parametrización por el arco, 106
- péndulo
 - esférico, 393
- péndulo de Foucault, 274
- peso, 223
- polares (coordenadas), 34
- potenciañ, 330
- primitiva, 69, 70
- producto
 - escalar, 456
 - mixto, 463
 - vectorial, 460
- prolongación analítica (principio de), 445
- pseudoesfera, 479
- radio de curvatura, 210
- regla de la cadena, 14, 304, 308
- reparametrización, 102
- Sarrus (regla de), 464
- segmento, 423
- seno/coseno lemniscáticos, 370
- sinodal (curva), 338
- sistema de referencia, 453
- Snell (ley de), 26
- subnormal, 33
 - polar, 34
- subtangente, 33
 - polar, 34
- tangente
 - recta, 3, 9
 - variedad, 282
 - vector, 204
- tautócrona (curva), 336
- Taylor (serie de), 443
- Teorema
 - de aproximación de Weierstrass, 95
 - de Arzelà, 139
 - de cambio de variable, 85

- de Cauchy, 28, 437
 - para triángulos, 436
- de Darboux, 51
- de Fermat, 22
- de L'Hôpital, 43–45
- de la convergencia monótona,
 - 158
- de la función inversa, 18, 309
- de Liouville, 443
- de Morera, 438
- de Rolle, 27
- de Schwarz, 301
- del valor medio, 28
- fundamental del álgebra, 444
- torsión, 211
- trabajo, 326
- tractriz, 477

- valor principal, 149
- vector, 454
- velocidad, 201
 - angular, 216, 219
 - sobre la trayectoria, 202