Resumen que puede usarse en el examen

Tema 2. Optimización sobre un conjunto convexo.

Min $f(x)$
s.a. $x \in X$ f continuamente diferenciable sobre $X \subseteq \mathbb{R}^n$, convexo, cerrado no vacío.

Propiedad 1
Sea C un convexo de $\mathbb{R}^n$ no vacío y $f : C \to \mathbb{R}$ una función convexa, entonces los conjuntos de nivel inferior $L_f(\alpha) = \{x \in C : f(x) \leq \alpha\}$ son convexos para cualquier escalar $\alpha \in \mathbb{R}$.

Propiedad 2
Dada una colección de convexos $\{C_i : i \in I\}$, su intersección $\cap_{i \in I} C_i$ es un convexo.

Condiciones de optimalidad.

Proposición
(a) Si x^* es un mínimo local de f sobre el convexo X, entonces
$\nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in X$
(b) Si además f es convexa sobre X, la condición anterior es suficiente para asegurar que x^* minimiza f sobre X.

Proyección sobre un convexo cerrado *X*.

Teorema
(a) Para cada $z \in \mathbb{R}^n$ existe un único $x^* \in X$ que minimiza $\|z - x\|$ sobre X. Éste vector es la proyección de z sobre X y se denota mediante $[z]^+$.
(b) Dado $z \in \mathbb{R}^n$, un vector $x^* \in X$ es la proyección de z sobre X, si y solo si: $(z - x^*)^T (x - x^*) \leq 0 \quad \forall x \in X$.
(c) La aplicación $f : \mathbb{R}^n \to \mathbb{R}^n$ definida por $f(x) = [x]^+$ es contínua y no expansiva.
(d) Si X es un subespacio, un vector $x^* \in X$ es la proyección de z si y solo si $(z - x^*)^T x = 0 \quad \forall x \in X$

Métodos de direcciones factibles

Definición. Dado x^k, una dirección factible en x^k es un vector $d^k \neq 0_n$ tal que $\exists \delta > 0 : \forall \alpha \in [0, \delta], \; x^k + \alpha d^k \in X$.

En estos métodos se comienza desde una solución factible x^0 y se genera una sucesión de vectores factibles $\{x^k\}$ como $x^{k+1} = x^k + \alpha^k d^k$ siendo d^k una dirección de descenso factible en x^k y α^k la longitud de salto adecuada.

Métodos de proyección del gradiente

El más sencillo genera una sucesión de iterados de la siguiente forma:
$x^{k+1} = x^k + \alpha^k(\overline{x}^k - x^k) \quad \alpha^k \in (0,1]$, siendo $\overline{x}^k = \left[x^k - s^k \nabla f(x^k)\right]^+$ y $s^k > 0, \forall k$
Para calcular la longitud de salto puede utilizarse la regla de minimización limitada o la regla de Armijo para direcciones factibles.

Suboptimización en variedades

Min *f(x)*
s.a. $a_j^T x \le b_j \quad j = 1,\ldots,r$

<u>Hipótesis:</u> Para cada $x \in X$ el conjunto de vectores $\{a_j : j \in A(x)\}$ es linealmente independiente, siendo $A(x) = \{j : a_j^T x = b_j\}$ (conjunto de índices de las restricciones activas)

<u>Esquema general</u>
Dado $x^k \in X$ se busca una dirección factible de descenso d^k en el subespacio $S(x^k) = \{d : a_j^T d = 0 \quad \forall j \in A(x^k)\}$

(a) Si existe dicha dirección $d^k \in S(x^k)$ factible de descenso, $x^{k+1} = x^k + \alpha^k d^k$ y α^k se calcula en el intervalo $\{\alpha > 0 : x^k + \alpha d^k \in X\}$

(b) Si no existe ninguna, x^k es estacionario sobre la variedad $x^k + S(x^k)$

- b1) Si x^k es estacionario sobre $X = \{x : a_j^T x \le b_j \quad j = 1,\ldots,r\}$ el algoritmo para.
- b2) Si x^k no es estacionario sobre *X*, puede relajarse una de las restricciones $\overline{j}$ y se obtiene una dirección factible de descenso en el subespacio

$$\overline{S}(x^k) = \{d : a_j^T d = 0 \quad \forall j \in A(x^k) \sim \{\overline{j}\}\}$$

Para determinar $d^k \in S(x^k)$ dirección factible de descenso se resuelven problemas cuadráticos de la forma:

(P1) Min $\nabla f(x^k)^T d + \frac{1}{2} d^T H^k d$

s.a. $a_j^T d = 0 \quad j \in A(x^k)$

con H^k matriz simétrica definida positiva.

La solución óptima de éste problema puede verse como la proyección a escala H^k del gradiente $\nabla f(x^k)$ sobre el subespacio $S(x^k)$, entonces $d^k = -(H^k)^{-1}(\nabla f(x^k) + (A^k)^T \mu)$, donde $\mu = -A^k (H^k)^{-1}(A^k)^T)^{-1} A^k (H^k)^{-1} \nabla f(x^k)$ y A^k es una matriz cuyas filas son los vectores $a_j \;\; j \in A(x^k)$.

Tenemos que: $d^k = 0_n \Leftrightarrow \nabla f(x^k) + (A^k)^T \mu = 0_n \Leftrightarrow \exists \mu_j,\, j \in A(x^k) : \nabla f(x^k) + \sum_{j \in A(x^k)} \mu_j a_j = 0_n$

Entonces x^k es estacionario sobre X si y solo si $\mu_j \geq 0, \;\; \forall j \in A(x^k)$.

Por tanto, si $d^k = 0_n$ y x^k no es estacionario, existirá algún índice $\bar{j}$ tal que $\mu_{\bar{j}} < 0$

Proposición

En las condiciones actuales $(\nabla f(x^k) + (A^k)^T \mu = 0_n, \mu_{\bar{j}} < 0)$, sea $\bar{d}^k$ la única solución óptima del problema :

(P2) Min $\nabla f(x^k)^T d + \frac{1}{2} d^T \bar{H}^k d$

s.a $\quad a_j^T d = 0 \quad j \in A(x^k) \sim \{\bar{j}\}$ $\qquad\qquad$ $\bar{H}^k$ matriz definida positiva

Entonces $\bar{d}^k$ es una dirección factible de descenso en x^k.

Programación Cuadrática

Min $\frac{1}{2} x^T Q x + c^T x$

s.a. $\quad a_j^T x \leq b_j \;\; j = 1, \ldots, r$

Las matrices H^k y $\bar{H}^k$ se toman iguales a Q.

Acerca del cálculo de la longitud de salto a lo largo de una dirección de descenso:

Si $d^k \neq 0_n$, comprobaremos si $x^k + d^k \in X$, en cuyo caso $x^{k+1} = x^k + d^k$, si $x^k + d^k \notin X$, entonces $x^{k+1} = x^k + \alpha^k d^k$, siendo $\alpha^k = \max\{\alpha > 0 : x^k + \alpha d^k \in X\}$, por tanto

$$\alpha^k = \min_i \left\{ \frac{\hat{b}_i}{\hat{d}_i} : \hat{d}_i > 0 \right\}$$

siendo $\hat{d}_j = a_j^T d^k \;\; j \notin A(x^k), \;\; \hat{b}_j = b_j - a_j^T x^k$.

¿Por que no puede suceder que $\hat{d}_j \leq 0 \quad \forall j \notin A(x^k)$?

Tema 3. Teoría de los multiplicadores de Lagrange.

Condiciones necesarias para restricciones de igualdad

Consideremos el problema:

(P) Min *f(x)* $f:\mathbb{R}^n \to \mathbb{R},\ \ h_i:\mathbb{R}^n \to \mathbb{R}$ continuamente diferenciables.
s.a. $h_i(x) = 0 \ \ i=1,...,m$

Teorema de los multiplicadores de Lagrange

Sea x^* un mínimo local de (P) y supongamos que los gradientes de las restricciones $\{\nabla h_i(x^*) : i=1,\dots,m\}$ son linealmente independientes.

(i) Entonces, existe un único vector $\lambda^* = (\lambda_1^*,\dots,\lambda_m^*)^T$ llamado vector de los multiplicadores de Lagrange, tal que $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0_n$

(ii) Si, además *f* y *h* son dos veces continuamente diferenciables, se tiene que:
$y^T(\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*))y \geq 0 \ \ \forall y \in V(x^*)$, siendo *V(x*)* el subespacio de las variaciones factibles de primer orden: $V(x^*) = \{y : \nabla h_i(x^*)^T y = 0 \ \ \text{i=1,...,m}\}$

Definición. Función lagrangiana

$L:\mathbb{R}^{n+m} \to \mathbb{R}$, definida como $L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)$

Condiciones suficientes para restricciones de igualdad

Proposición.
Supongamos que *f* y *h* son dos veces continuamente diferenciables. Sean $x^* \in \mathbb{R}^n$ y $\lambda^* \in \mathbb{R}^m$, que satisfacen:

$$\nabla_x L(x^*,\lambda^*) = 0_n \qquad \nabla_\lambda L(x^*,\lambda^*) = 0_m \qquad (1)$$

$$y^T \nabla^2_{xx} L(x^*) y > 0 \ \ \forall y \neq 0_n \ \ \text{tal que } \nabla h(x^*)^T y = 0 \qquad (2)$$

Entonces x^* es un mínimo local estricto de *f* restringido a $h(x) = 0_m$. De hecho, existen escalares $\gamma>0$ y $\varepsilon>0$ tales que

$$f(x) \geq f(x^*) + \frac{\gamma}{2}\|x - x^*\|^2 \quad \forall x : h(x) = 0_m \ y \ \|x - x^*\| < \varepsilon$$

La función lagrangiana aumentada

Dado cualquier escalar c, se define la función lagrangiana aumentada

$$L_c(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \frac{c}{2}\|h(x)\|^2$$

Restricciones de desigualdad: condiciones de K-K-T

Consideremos el problema más general:

(ICP) Min $f(x)$ $\quad$ f, h_i y g_j funciones continuamente diferenciables.

s.a. $h_i(x) = 0 \quad i=1,\ldots,m$

$g_j(x) \le 0 \quad j=1,\ldots,r$

- Dada una solución factible x el conjunto de restricciones de desigualdad activas se denota $A(x) = \{j : g_j(x) = 0\}$.
- Una solución factible es <u>regular</u> si el conjunto de vectores siguiente es linealmente independiente: $\{\nabla h_i(x), i=1,\ldots,m, \nabla g_j(x),\ j \in A(x)\}$
- Para (ICP) la función lagrangiana se define: $L(x,\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x)$

Condiciones necesarias de optimalidad

Sea x^* un mínimo local de (ICP) que es regular.

(1) Entonces, existe un único vector (λ^*,μ^*), siendo $\lambda^* = (\lambda_1^*,\ldots,\lambda_m^*)^T$,

$\mu^* = (\mu_1^*,\ldots,\mu_r^*)^T$ llamado vector de los multiplicadores de Lagrange, tal que

$\nabla_x L(x^*,\lambda^*,\mu^*) = 0_n \quad \mu_j \ge 0 \quad j=1,\ldots,r$ y $\mu_j = 0$ si $j \notin A(x^*)$

[(x^*,λ^*,μ^*) es un punto de Karush, Kuhn y Tucker para (ICP)]

(2) Si, además f ,g y h son dos veces continuamente diferenciables, se tiene que:

$y^T \nabla_{xx}^2 L(x^*,\lambda^*,\mu^*) y \ge 0 \quad \forall y \in V(x^*)$ siendo

$V(x^*) = \{y : \nabla h_i(x^*)^T y = 0 \ \text{ i=1,...,m}, \nabla g_j(x^*)^T y = 0,\ j \in A(x^*)\}$

Condiciones suficientes de optimalidad

Sean $x^* \in \mathbb{R}^n, \lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^r$ tales que:

$h(x^*) = 0_m, g(x^*) \le 0_r$

$\nabla_x L(x^*,\lambda^*,\mu^*) = 0_n \quad \mu_j \ge 0 \quad j=1,\ldots,r$ y $\mu_j = 0$ si $j \notin A(x^*)$

y también verifican que

$y^T \nabla_{xx}^2 L(x^*,\lambda^*,\mu^*) y > 0 \quad \forall y \in V(x^*) \quad y \ne 0_n$ siendo

$V(x^*) = \{y : \nabla h_i(x^*)^T y = 0 \ \text{ i=1,...,m}, \nabla g_j(x^*)^T y = 0,\ j \in A(x^*)\}$.

Supongamos que $\mu_j > 0 \quad \forall j \in A(x^*)$, entonces x^* es mínimo local estricto de (ICP).

Tema 4. Métodos basados en el uso de las penalizaciones.

1. Métodos de barrera

(P) Min $f(x)$ — f y g_j continuas

s.a. $g_j(x) \leq 0$ j=1,...,r

$x \in X$ $X \subseteq \mathbb{R}^n$ cerrado

Sucesión de parámetros $\{\varepsilon^k\}$ que verifican $0 < \varepsilon^{k+1} < \varepsilon^k$, k = 0,1, ..., con $\varepsilon^k \to 0$

Hallar $x(\varepsilon^k) := \arg\min_{x \in S} \{ f(x) + \varepsilon^k B(x) \}$ k = 0,1, ..., siendo

$S = \{ x \in X : g_j(x) < 0,\ j = 1, ..., r \}$ y $B(x)$ la función barrera que está definida en $S \neq \emptyset$, es continua y tiende a ∞ si alguna de las restricciones se aproxima a 0. Las más habituales:

$B(x) = -\sum_{j=1}^{r} \ln(-g_j(x))$ y $B(x) = -\sum_{j=1}^{r} \frac{1}{g_j(x)}$

Proposición.
Cada punto límite de la sucesión $\{x^k\}_{k \geq 0}$ generada mediante un método de barrera es un mínimo global de (P)

2. Métodos de penalización y de los multiplicadores

(P) Min $f(x)$

s.a. $h_i(x) \leq 0$ i=1,...,m

$x \in X$ $X \subseteq \mathbb{R}^n$

Lagrangiano aumentado: $L_c(x, \lambda) = f(x) + \lambda^T h(x) + \frac{c}{2} \|h(x)\|^2$, c>0

Proposición.
Supongamos que f y h son continuas y X un conjunto cerrado, también suponemos que $\{ x \in X : h(x) = 0_m \} \neq \emptyset$. Para k=0, 1,..., sea x^k un mínimo global del problema

Min $L_{c^k}(x, \lambda^k)$

s.a. $x \in X$

siendo $\{\lambda^k\}$ una sucesión acotada, $0 < c^k < c^{k+1}$ ∀k y $\{c^k\} \to \infty$.

Entonces cada punto límite de la sucesión $\{x^k\}$ es un mínimo global de (P).

Esquema numérico del SUMT

Inicializar: Elegir ε>0, $\beta\in[4, 10]$,, c^0>0 y $x^0\in \mathbb{R}^n$ .Hacer k = 0

Etapa1: Empezando en x^k resolver el problema

Min $L_{c^k}(x,0) = f(x) + \frac{c^k}{2}\|h(x)\|^2$

Llamar x^{k+1} a la solución óptima del problema irrestringido

Etapa 2: Si $\frac{c^k}{2}\|h(x^{k+1})\|^2 < \varepsilon$, entonces parar.

En otro caso hacer $c^{k+1} = \beta c^k$, k = k+1 y repetir la etapa 1.

Esquema numérico del método de los multiplicadores (ALAG)

Inicializar: Elegir ε>0, $\beta\in[5, 10]$, c^0>0, $\lambda^0\in \mathbb{R}^m$. Tomar $x^0 = 0_n$ Hacer k = 0

Minimización de la función de penalización

Tomando como solución inicial x^k, resolver el problema Min $L_{c^k}(x,\lambda^k)$ y llamar x^{k+1} a la solución óptima obtenida

Si $\|h(x^{k+1})\| < \varepsilon$, parar y $x^{k+1} = x^*$ es mínimo local del problema restringido.

Si $\|h(x^{k+1})\| \geq 0.25\|h(x^k)\|$, reemplazar c^k por $c^{k+1} = \beta c^k$, $\lambda^{k+1} = \lambda^k$ y repetir ésta etapa.

Si $\|h(x^{k+1})\| < 0.25\|h(x^k)\|$, actualizar los multiplicadores $\lambda^{k+1} = \lambda^k + c^k h(x^k)$, y $c^{k+1} = c^k$ y repetir ésta etapa. $L_{c^k}(x,\lambda^k) = f(x) + (\lambda^k)^T h(x) + \frac{c^k}{2}\|h(x)\|^2$

3. Métodos de Lagrange-Newton y programación cuadrática secuencial.

Método tipo Newton para problemas con restricciones de igualdad

Resolver el sistema lagrangiano:

$$\left.\begin{aligned} \nabla f(x) + \nabla h(x)\lambda &= 0_n \\ h(x) &= 0_m \end{aligned}\right\} \Leftrightarrow \nabla L(x,\lambda) = 0_{n+m}$$

mediante el método de Newton : $x^{k+1} = x^k + \Delta x^k$ y $\lambda^{k+1} = \lambda^k + \Delta\lambda^k$, donde (x^k,λ^k) se obtiene resolviendo el sistema de ecuaciones:

$$\nabla^2 L(x^k,\lambda^k)\begin{pmatrix}\Delta x^k \\ \Delta\lambda^k\end{pmatrix} = -\nabla L(x^k,\lambda^k) \qquad (1)$$

Bajo ciertas condiciones sobre x^* (mínimo local regular) se puede asegurar que, en un entorno suyo, los puntos generados mediante (1) estan bien definidos . En éste caso

$$\nabla L(x^k,\lambda^k) = \begin{pmatrix}\nabla_x L(x^k,\lambda^k) \\ h(x^k)\end{pmatrix} \text{y } \nabla^2 L(x^k,\lambda^k) = \begin{pmatrix} H^k & N^k \\ (N^k)^T & 0_{mxm}\end{pmatrix}$$

siendo

$$H^k = \nabla^2_{xx} L(x^k,\lambda^k) = \nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) \text{ y } N^k = \nabla h(x^k) = \left(\nabla h_1(x^k),\ldots,\nabla h_m(x^k)\right)$$

Si H^k es invertible y el rango de N^k es m, podemos obtener una expresión más explícita para la iteración de Newton:

$$\lambda^{k+1} = \left((N^k)^T (H^k)^{-1} (N^k)\right)^{-1} \left(h(x^k) - (N^k)^T (H^k)^{-1} \nabla f(x^k)\right)$$

$$x^{k+1} = x^k - (H^k)^{-1} \nabla_x L(x^k, \lambda^{k+1})$$

<u>Programación Cuadrática Secuencial (SQP)</u>

Conocido (x^k, λ^k) se determina $(\Delta x^k, \lambda^{k+1})$ resolviendo el problema cuadrático:

$$\text{PC}(x^k, \lambda^k) \quad \text{Min} \ \nabla f(x^k)^T \Delta x + \frac{1}{2} \Delta x^T H^k \Delta x$$

$$\text{s.a.} \quad h(x^k) + (N^k)^T \Delta x = 0_m$$

Extensión para incluir <u>restricciones de desigualdad</u>:

$$\text{PC}(x^k, \lambda^k, \mu^k) \quad \text{Min} \ \nabla f(x^k)^T \Delta x + \frac{1}{2} \Delta x^T H^k \Delta x$$

$$\begin{aligned} \text{s.a.} \quad & h_i(x^k) + \nabla h_i(x^k)^T \Delta x = 0 \quad i = 1, \dots, m \\ & g_j(x^k) + \nabla g_j(x^k)^T \Delta x \le 0 \quad \text{j} = 1, \dots, r \end{aligned}$$

siendo $H^k = \nabla_{xx}^2 L(x^k, \lambda^k) = \nabla^2 f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla^2 g_j(x^k)$