Resumen que puede usarse en el examen

Tema 1. Optimización Irrestringida.

Condiciones necesarias y suficientes de optimalidad.

Proposición (C. Necesarias)
Sea x^* un mínimo local irrestringido de $f:\mathbb{R}^n \to \mathbb{R}$ y supongamos que *f* es continuamente diferenciable sobre un abierto S que contiene a x^*, entonces $\nabla f(x^*) = 0_n$. Si, además *f* es dos veces continuamente diferenciable en S, entonces $\nabla^2 f(x^*)$ es semidefinida positiva.

Proposición (Caracterización de las funciones convexas diferenciables)
Sea $C \subseteq \mathbb{R}^n$ un convexo y $f:C \to \mathbb{R}$ una función diferenciable sobre *C*.
(a) La función *f* es convexa sii $f(z) \geq f(x) + (z-x)^T \nabla f(x) \quad \forall x, z \in C$
(b) Si la desigualdad anterior es estricta cuando $x \neq z$, entonces *f* es estrictamente convexa.

Proposición (función objetivo convexa)
Sea $f:C \to \mathbb{R}$ una función convexa sobre el convexo *C*,
(a) Cualquier mínimo local de *f* sobre *C* es también mínimo global sobre *C*. Si además *f* es estrictamente convexa, entonces como mucho existe un mínimo global de *f*.
(b) Si *f* es convexa y *C* abierto, entonces $\nabla f(x^*) = 0_n$ es una condición necesaria y suficiente para que $x^* \in C$ sea mínimo global de f sobre *C*.

Proposición (C. Suficiente)
Sea $f:\mathbb{R}^n \to \mathbb{R}$ dos veces continuamente diferenciable en un abierto C.
Supongamos que $x^* \in C$ satisface que $\nabla f(x^*) = 0_n$ y $\nabla^2 f(x^*)$ es una matriz definida positiva. Entonces x^* es un mínimo local irrestringido de *f*. En particular se tiene que:
$\exists\ \gamma > 0, \varepsilon > 0$ tales que $f(x) \geq f(x^*) + \dfrac{\gamma}{2}\|x - x^*\|^2 \ \ \forall x : \|x - x^*\| < \varepsilon$

Proposición
Sea $C \subseteq \mathbb{R}^n$ un convexo y $f:\mathbb{R}^n \to \mathbb{R}$ dos veces continuamente diferenciable en C. Sea *Q* una matriz real *n* x *n* simétrica,
(a) Si $\nabla^2 f(x)$ es semidefinida (definida) positiva $\forall x \in C$, entonces *f* es convexa (estrictamente convexa).
(b) Si $C = \mathbb{R}^n$ y *f* es convexa, entonces $\nabla^2 f(x)$ es semidefinida positiva $\forall x$
(c) La función $f(x) = \dfrac{1}{2} x^T Q x$ es convexa sii *Q* es semidefinida positiva.

Métodos de descenso basados en el gradiente

Dado x^k, iterado k-ésimo, si $\nabla f(x^k) \neq 0_n$, definimos $x^{k+1} = x^k + \alpha^k d^k$ siendo d^k una dirección tal que $\nabla f(x^k)^T . d^k < 0$ y α^k la longitud de salto adecuada.

Las direcciones de descenso

En muchos casos $d^k = -D^k \nabla f(x^k)$, siendo D^k una matriz simétrica definida positiva:

Descenso más rápido : $D^k = I_{n \times n} \quad \forall k$

Newton : $D^k = \left(\nabla^2 f(x^k)\right)^{-1} \quad \forall k \geq 0$

Newton modificado: En lugar de calcular cada vez la inversa del hessiano, sólo se hace cada cierto número de iteraciones.

Esquema algorítmico

Inicialización: Elegir $\varepsilon>0$ y $x^1 \in R^n$. Hacer k=1 e ir al paso 1.

Paso 1. Evaluar $\nabla f(x^k)$. Si $\| \nabla f(x^k) \| < \varepsilon \rightarrow$STOP (nos quedamos con x^k). En otro caso, hacer $d^k = -D^k \nabla f(x^k)$ e ir a 2.

Paso 2. Evaluar la longitud del desplazamiento $\alpha_k > 0$ e ir a 3.

Paso 3. Hacer $x^{k+1} = x^k + \alpha_k d^k$. Hacer k=k+1 y volver a 1.

Selección de la amplitud de salto

Regla de minimización (búsqueda lineal exacta)

Elegimos $\alpha^k \in Arg \min_{\alpha \geq 0} f(x^k + \alpha d^k)$

Regla de minimización limitada

Elegimos $\alpha^k \in Arg \min_{\alpha \in [0,s]} f(x^k + \alpha d^k)$ para un escalar fijo s>0

Estas reglas se implementan con ayuda de algoritmos de búsqueda lineal unidimensional (ver apéndices A1 y A2)

Reducción sucesiva

1 Búsqueda lineal mediante la **Regla de Armijo**

Inicialización: Elegir s, $\beta \in (0.1, 0.5)$, $\sigma \in [10^{-5}, 10^{-1}]$. Sea $x \in R^n$ arbitrario y d una dirección de descenso en x. Sea m=0, ir a 1.

Paso 1. Evaluar G(m)= $f(x) - f(x + \beta^m s d)$ y g(m)=$-\sigma \beta^m s \nabla f(x)^T d$. Si G(m)≥g(m) $\rightarrow$STOP y $\alpha = \beta^m s$. En otro caso, hacer m=m+1 y repetir el paso 1.

2 **Regla de Goldstein**

Se fija un escalar $\sigma \in \left(0, \frac{1}{2}\right)$ y se elige α^k de modo que se satisfaga la relación:

$$\sigma < \frac{f(x^k + \alpha^k d^k) - f(x^k)}{\alpha^k \nabla f(x^k)^T d^k} \leq 1 - \sigma$$

3 **Reglas de Wolfe**

Se elige α^k de modo que se satisfagan las relaciones:

$$f(x^k + \alpha^k d^k) \leq f(x^k) + c_1 \alpha^k \nabla f(x^k)^T d^k$$
$$\nabla f(x^k + \alpha^k d^k)^T d^k \geq c_2 \nabla f(x^k)^T d^k$$
$$\text{siendo } 0 < c_1 < c_2 < 1$$

Resultados de convergencia

"Los métodos de descenso nos conducen hacia los puntos estacionarios más cercanos"

Si $\left\{x^k\right\}_{k \geq 0}$ es la sucesión de iterados obtenida al aplicar uno de estos métodos, la existencia de límite de ésta sucesión está asegurada si el conjunto de nivel inferior $\left\{x : f(x) \leq f(x^0)\right\}$ está acotado, siendo x^0 la solución inicial.

Para asegurar que el límite es un punto estacionario hay que añadir otras condiciones técnicas, por ejemplo:

C1. Sea $d^k = -D^k \nabla f(x^k)$, supongamos que los valores propios de la matriz simétrica definida positiva D^k están acotados superior e inferiormente por el cero. Es decir, existen dos escalares positivos c_1 y c_2 tales que: $c_1 \|z\|^2 \leq z^T D^k z \leq c_2 \|z\|^2 \quad \forall z \in \mathbb{R}^n, \forall k \geq 0$

C2. La sucesión de direcciones de descenso $\left\{d^k\right\}_{k \geq 0}$ es gradiente afín para $\left\{x^k\right\}_{k \geq 0}$ si se satisface la condición siguiente: "Para cualquier subsucesión $\left\{x^k\right\}_{k \in K}$ que converge a un punto no estacionario, la correspondiente subsucesión $\left\{d^k\right\}_{k \in K}$ está acotada y cumple que $\lim_{k \to \infty} \sup_{k \in K} \nabla f(x^k)^T d^k < 0$.

Proposición

Sea $\left\{x^k\right\}_{k \geq 0}$ la sucesión generada por un método basado en el gradiente. Supongamos que $\left\{d^k\right\}_{k \geq 0}$ es gradiente afín y que α^k se ha determinado mediante la regla de Armijo o una búsqueda lineal exacta. Entonces cada punto límite de $\left\{x^k\right\}_{k \geq 0}$ es estacionario.

Tasa de convergencia

Análisis local

La tasa se evalúa en términos de una función de error $e:\mathbb{R}^n \to \mathbb{R}$, $e(x) \geq 0$. Habitualmente $e_1(x) = \left\| x - x^* \right\|$ o $e_2(x) = \left| f(x) - f(x^*) \right|$.

Se dice que $\left\{ e(x^k) \right\}_{k\geq 0}$ converge <u>linealmente</u> o geométricamente si $\exists q > 0$ y $\exists \beta \in (0,1)$: $e(x^k) \leq q\beta^k \ \forall k$. Lo que se obtiene si para algún $\beta \in (0,1)$ se cumple $\lim_{k\to\infty} \sup \frac{e(x^{k+1})}{e(x^k)} \leq \beta$.

Si para cada $\beta \in (0,1)$, $\exists q > 0$: $e(x^k) \leq q\beta^k \ \forall k$, diremos que $\left\{ e(x^k) \right\}_{k\geq 0}$ converge superlinealmente. Lo que se obtiene, en particular, si $\lim_{k\to\infty} \sup \frac{e(x^{k+1})}{e(x^k)} = 0$

Proposición.

Sea $f(x) = \frac{1}{2} x^T Q x$ donde Q es simétrica definida positiva. Consideremos el método de descenso más rápido, eligiendo α^k mediante la regla de minimización, entonces $\forall k$:
$f(x^{k+1}) \leq \left(\frac{M-m}{M+m} \right)^2 f(x^k)$ (es decir la convergencia de la sucesión de errores es lineal), siendo m y M los valores propios de Q menor y mayor respectivamente.

Métodos de direcciones conjugadas.

Dada una matriz Q n x n definida positiva, diremos que los vectores no nulos $d^1,\dots,d^k$ son Q-conjugados si $(d^i)^T Q d^j = 0 \ \ \forall i, j \ \ i \neq j$.

Dado un conjunto de vectores Q-conjugados $\left\{ d^0, d^1, \dots, d^{n-1} \right\}$, el método de direcciones conjugadas para <u>minimizar una función cuadrática</u> $f(x) = \frac{1}{2} x^T Q x - b^T x$, genera una sucesión de iterados $x^{k+1} = x^k + \alpha^k d^k \ \ k = 0,1,\dots,n-1$, siendo x^0 una solución inicial arbitraria y $\alpha^k = Arg \min_{\alpha} f(x^k + \alpha d^k)$.

Propiedad fundamental. Los sucesivos iterados minimizan *f* sobre la variedad lineal generada por las direcciones conjugadas. En particular, para cada *k*, x^{k+1} minimiza *f* sobre $M^k = L + x^0$, siendo L el subespacio vectorial generado por $\left\{ d^0, d^1, \dots, d^k \right\}$, es decir

$x^{k+1} = Arg \min_{x\in M^k} f(x)$, siendo $M^k = \left\{ x : x = x^0 + v \text{ donde } v = \sum_{j=0}^{k} \mu_j d^j \ \ \mu_j \in \mathbb{R} \ \ 1 \leq j \leq k \right\}$

Método del gradiente conjugado de Fletcher y Reeves.

Se obtiene aplicando el procedimiento de Gram-Schmidt a los vectores gradiente cambiados de signo. El método genera el iterado k+1 como: $x^{k+1} = x^k + \alpha^k d^k$, $\alpha^k = \arg\min f(x^k + \alpha d^k)$, si denotamos por $g_k = \nabla f(x^k)$, las direcciones vienen dadas por $d^0 = -g^0$, $d^k = -g_k + \beta_k d^{k-1}$ k=1,...,n con $\beta_k = \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$.

Fórmula de Polak-Ribiere

$$\beta_k = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}$$

Métodos casi-Newton

$x^{k+1} = x^k + \alpha^k d^k$ y $d^k = -D^k \nabla f(x^k)$, D^k definida positiva se va ajustando iteración tras iteración intentando que d^k se parezca a la dirección de Newton.

Fórmula de actualización

Llamamos $p_k = x^{k+1} - x^k$ y $q_k = \nabla f(x^{k+1}) - \nabla f(x^k)$

D^0 una matriz definida positiva

$$D^{k+1} = D^k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{D^k q_k q_k^T D^k}{q_k^T D^k q_k} + \xi^k \tau_k v_k v_k^T$$

donde $v_k = \dfrac{p_k}{p_k^T q_k} - \dfrac{D^k q^k}{\tau_k}$ y $\tau_k = q_k^T D^k q_k$ y $0 \le \xi^k \le 1 \ \ \forall k$

Si $\xi^k = 0 \ \ \forall k$, tenemos el método de Davidon- Fletcher- Powell

Si $\xi^k = 1 \ \ \forall k$, tenemos el método de Broyden- Fletcher- Goldfarb- Shanno

Proposición

Si D^k es una matriz definida positiva y elegimos la amplitud de salto α^k de manera que x^{k+1} satisface que $\nabla f(x^k)^T d^k < \nabla f(x^{k+1})^T d^k$, entonces D^{k+1} es definida positiva.

Proposición

Sean $\{x_k\},\{d_k\}$ y $\{D_k\}$ las sucesiones generadas por un algoritmo cuasi-Newton aplicado a la minimización de la función $f(x) = \dfrac{1}{2} x^T Q x - b^T x$ donde Q es simétrica definida positiva, en el que $\alpha_k = \arg\min_\alpha f(x^k + \alpha d^k)$.

Supongamos que ninguno de los vectores $x^0, x^1, ..., x^{n-1}$ es solución óptima, entonces:

a) Los vectores $\{d^0, d^1, ..., d^{n-1}\}$ son Q-conjugados.

b) $D^n = Q^{-1}$

APÉNDICE A1:

Método de Newton para encontrar los ceros de una función unidimensional ($g(\lambda)=0$)

Inicialización: Elegir $\varepsilon>0$ y $\lambda_1 \in$ R. Hacer k=1 e ir al paso 1.

Paso 1. Evaluar $g(\lambda_k)$ y $g'(\lambda_k) \neq 0$. Hacer $\lambda_{k+1} = \lambda_k - \dfrac{g(\lambda_k)}{g'(\lambda_k)}$ ir a 2.

Paso 2. Si $|\lambda_{k+1} - \lambda_k| < \varepsilon \rightarrow$STOP (nos quedamos con λ_{k+1}). En otro caso, hacer k=k+1 y volver a 1.

APÉNDICE A2:

Método de bisección para encontrar los ceros de una función unidimensional ($g(\lambda)=0$)

Inicialización: Elegir $\varepsilon>0$, a_1 , $b_1 \in$ R $a_1 < b_1$ tales que $g(a_1)$. $g(b_1) < 0$. Hacer k=1 e ir a 1.

Paso 1. Evaluar $\lambda_k = \dfrac{a_k + b_k}{2}$. Si $|b_k - a_k| < \varepsilon$. STOP (nos quedamos con λ_k) En otro caso evaluar $g(\lambda_k) \neq 0$, si $g(a_k)$. $g(\lambda_k) > 0$ ir a 2, si no ir a 3.

Paso 2. Hacer $a_{k+1} = \lambda_k$ y $b_{k+1} = b_k$ Hacer k= k+1 y volver a 1.

Paso 3. Hacer $a_{k+1} = a_k$ y $b_{k+1} = \lambda_k$ Hacer k= k+1 y volver a 1.