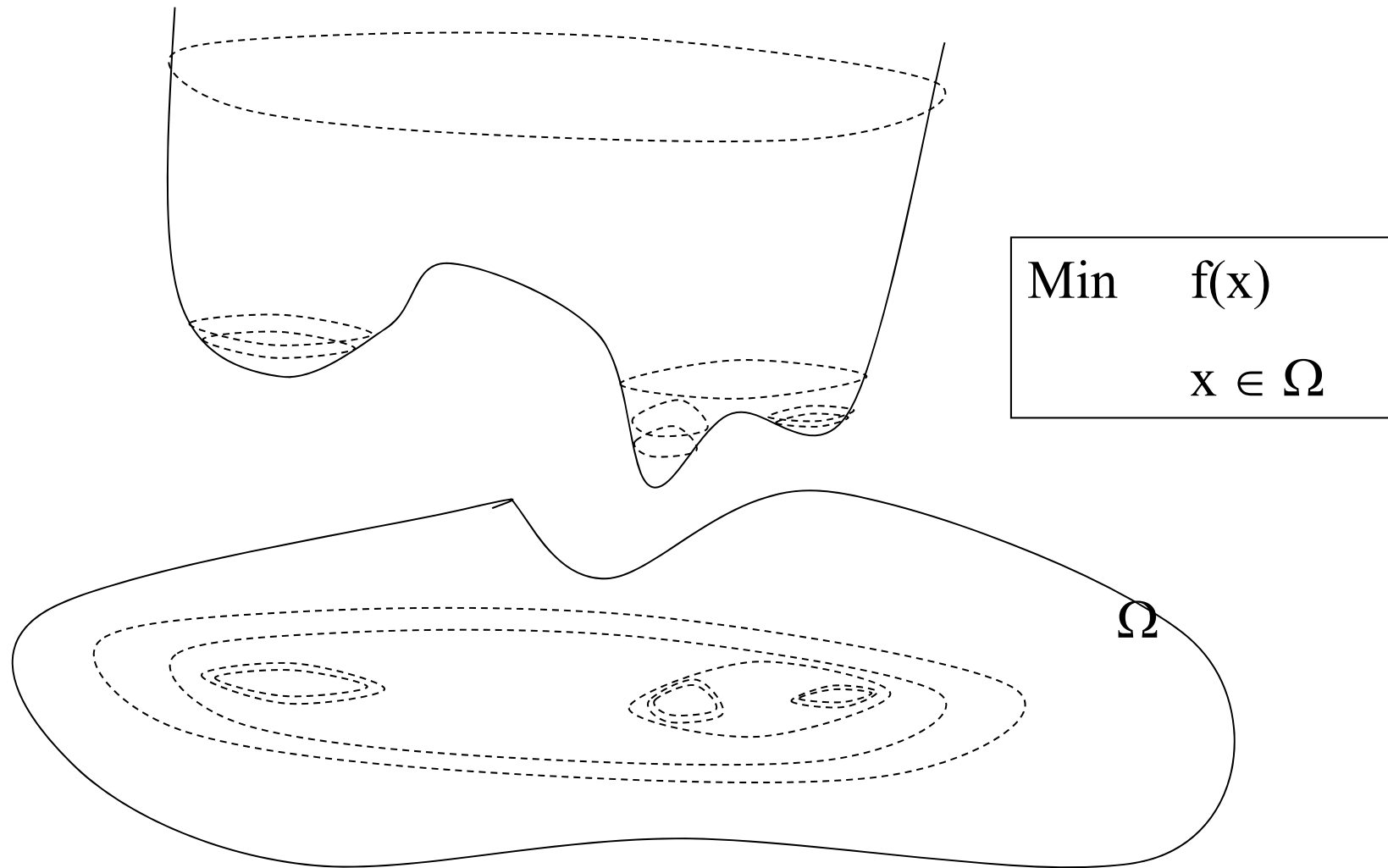
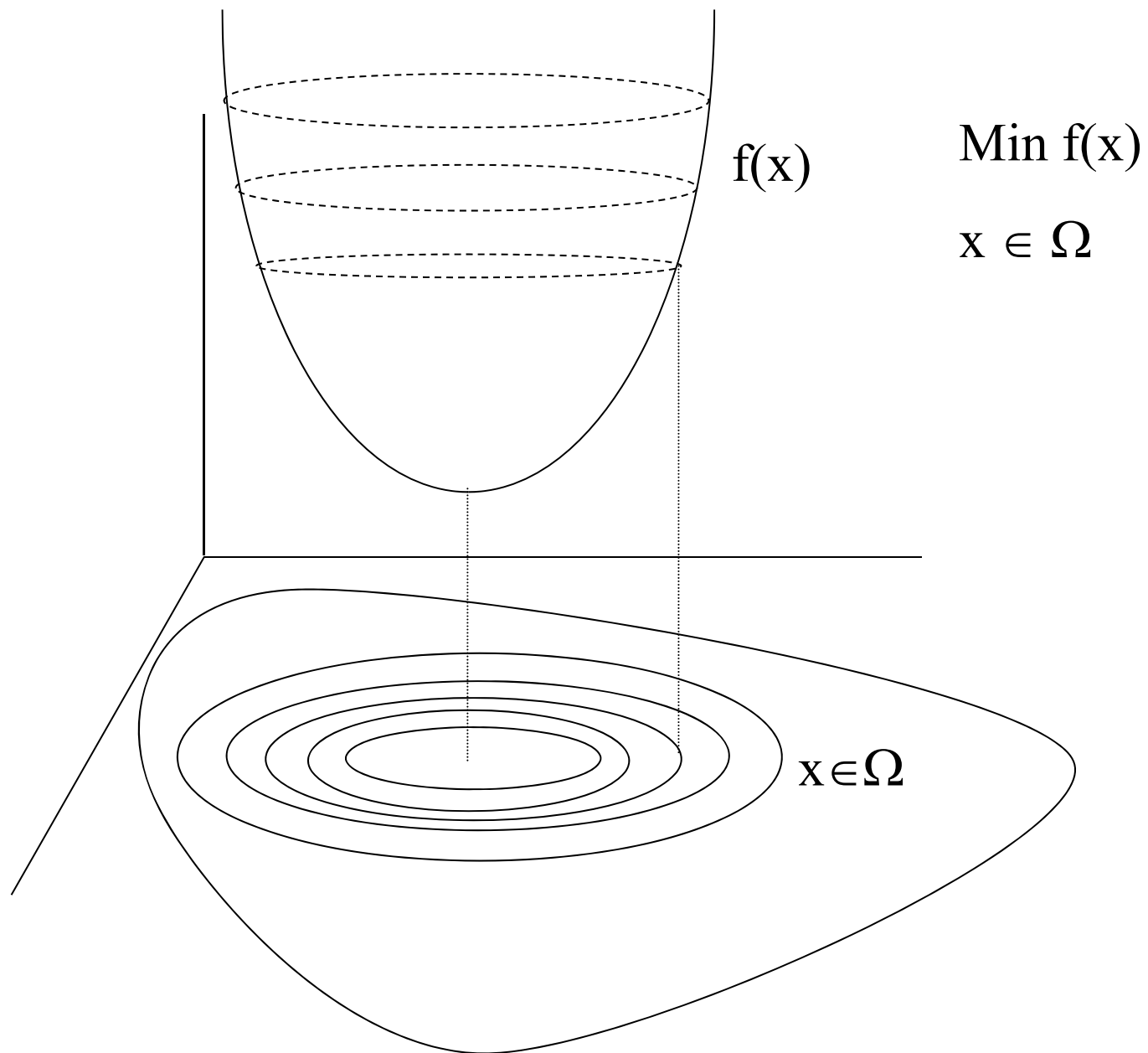


Optimización (No lineal)

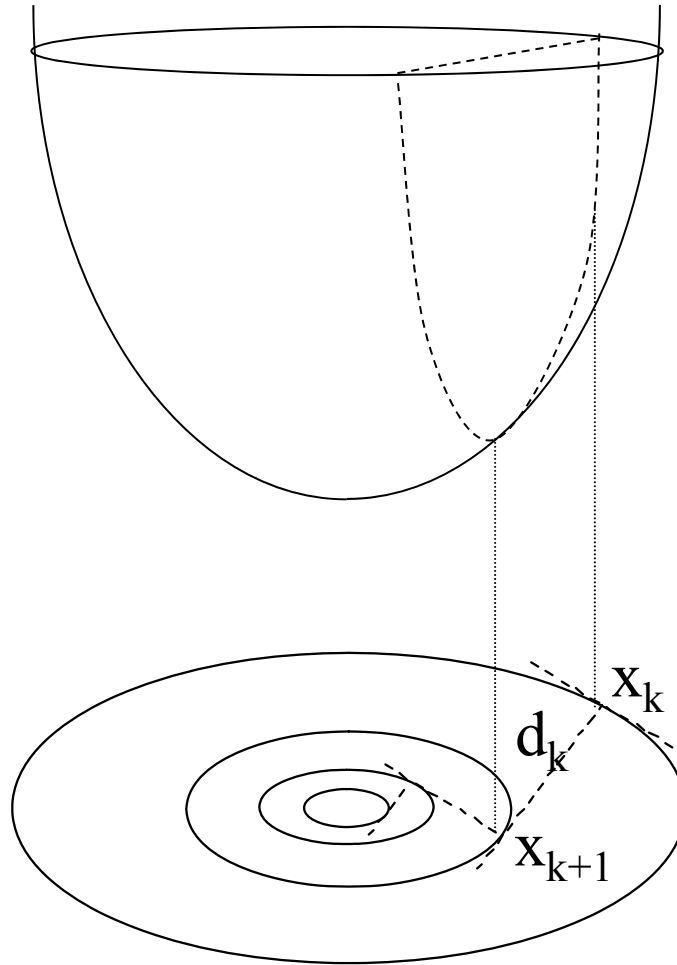




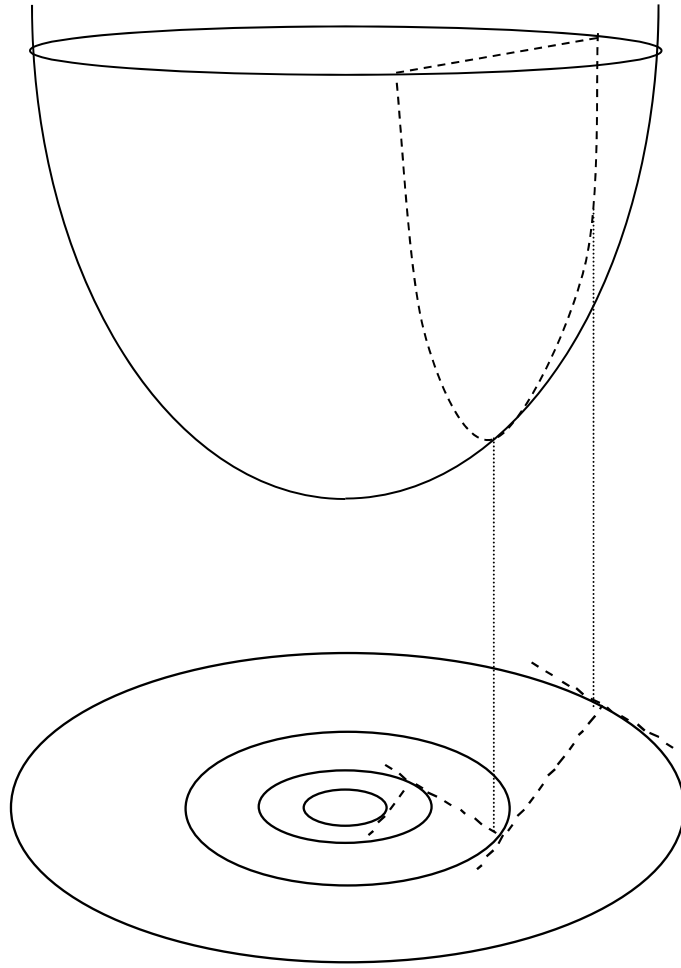
Métodos iterativos: $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$

- Criterio de selección de dirección de movimiento: \mathbf{d}_k
- Criterio de selección de longitud de paso: λ_k
(Exploración lineal)

Métodos iterativos: $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$



Una vez seleccionada la dirección de movimiento, el problema de obtener la “mejor” longitud de paso es un problema de optimización unidimensional



Una vez seleccionada la dirección de movimiento, el problema de obtener la “mejor” longitud de paso es un problema de optimización unidimensional

$$\text{Métodos iterativos: } \mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$$

- Criterio de selección de dirección de movimiento: \mathbf{d}_k

Caso no restringido: (\mathbf{d}_k de descenso)

- ▶ Método del gradiente $\mathbf{d}_k = -\nabla f(\mathbf{x})$
- ▶ Método de Newton $\mathbf{d}_k = -[\nabla^2 f(\mathbf{x})]^{-1} \nabla f(\mathbf{x})$

Caso Restringido: (\mathbf{d}_k factible y de descenso)

- ▶ Gradiente Proyectado

- Criterio de selección de longitud de paso: λ_k
(Exploración lineal)

Clases de problemas no lineales: Problemas sin restricciones: $\Omega = \mathbb{R}^n$.

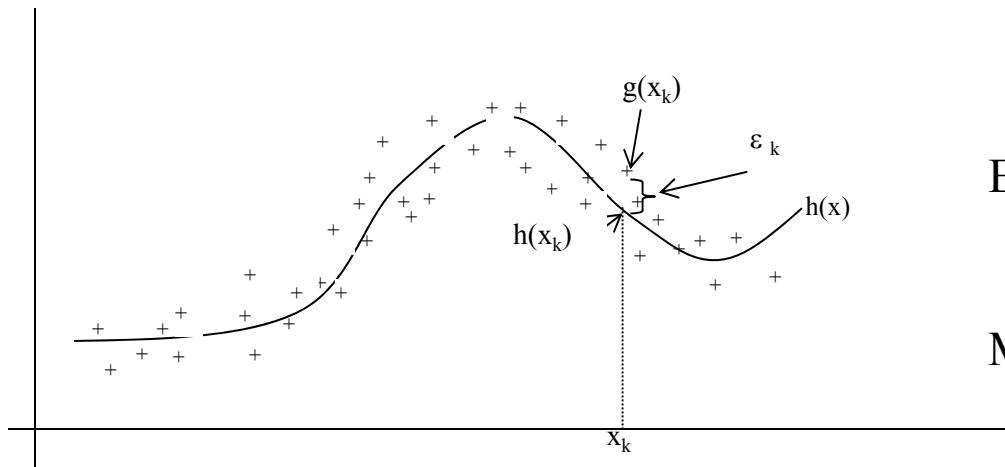
Ejemplo: Problema de aproximación. Hallar la estimación de una función $y=g(x)$

Valores observados

x	x_1	x_2	\dots	x_i	\dots	x_m
$g(x)$	$g(x_1)$	$g(x_2)$	\dots	$g(x_i)$	\dots	$g(x_m)$

Hipótesis : Aproximación de la función mediante un polinomio de grado $n < m$.

$$h(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$



Error aproximación: $\epsilon_k = g(x_k) - h(x_k)$

Mejor aproximación: $\text{Min} \sum_{k=1}^m \epsilon_k^2$

Problemas con restricciones: $\Omega \subset \mathbb{R}^n$.

Ejemplo: Una región está dividida en dos zonas, la 1 y la 2. Cada día los residentes en la zona 1 se desplazan hasta la zona 2, lo que genera un flujo de 25 habitantes desde el origen 1 hasta el destino 2, usando los posibles itinerarios en la red

El objetivo de cada viajero es minimizar su tiempo de viaje. El tiempo que tarda en recorrerse el arco es

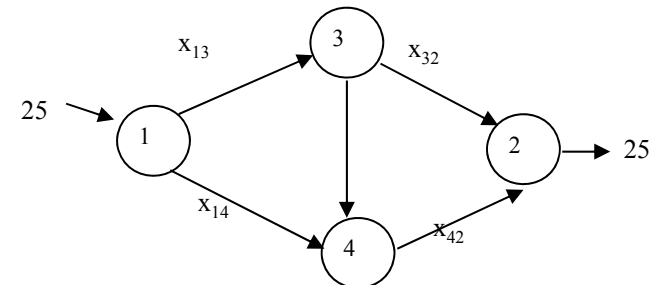
$$\text{Arco (1,3): } x_{13} \left[1 - \frac{x_{13}}{30} \right]^{-1}$$

$$\text{Arco (1,4): } 5 + 0.5 \cdot x_{14} \left[1 - \frac{x_{14}}{10} \right]^{-1}$$

$$\text{Arco (3,4): } 1 + x_{34} \left[1 - \frac{x_{34}}{10} \right]^{-1}$$

$$\text{Arco (3,2): } x_{32} \left[1 - \frac{x_{32}}{30} \right]^{-1}$$

$$\text{Arco (4, 2): } 5 + 0.1 \cdot x_{42} \left[1 - \frac{x_{42}}{30} \right]^{-1}$$



El flujo que minimiza el tiempo total de viaje :

$$\text{Min} \sum_{(i,j)} t_{ij}(x_{ij}) x_{ij}$$

sometido a condiciones de equilibrio del flujo.

Problema no lineal de flujos en redes:

Las ecuaciones de conservación de flujo son lineales, pero la función objetivo no lo es.

Existencia y caracterización de soluciones:

Teorema de Weierstrass (existencia): $f(x)$ continua, en Ω compacto $\Rightarrow f$ tiene mínimo en Ω
($\exists x^* \in \Omega$ tal que $f(x) \geq f(x^*) \forall x \in \Omega$)

Tª Weierstrass: cuándo existe el mínimo de una función en un dominio.

Pero... cuando se cumplan las condiciones de existencia, queremos **caracterizar** las soluciones y **saber como encontrarlas**.

Definición:

1. $x^* \in \Omega$ es un mínimo relativo (ó local) si $\exists \varepsilon > 0$ t.q. $f(x) \geq f(x^*) \forall x \in \Omega$ t.q. $|x - x^*| \leq \varepsilon$.
2. $x^* \in \Omega$ es un mínimo global si $f(x) \geq f(x^*) \forall x \in \Omega$.

Observaciones:

1. En general, sólo podremos obtener mínimos locales.
2. Si $f(x)$ posee ciertas propiedades (p.e. convexidad) en ocasiones podremos caracterizar los óptimos globales y conseguirlos.

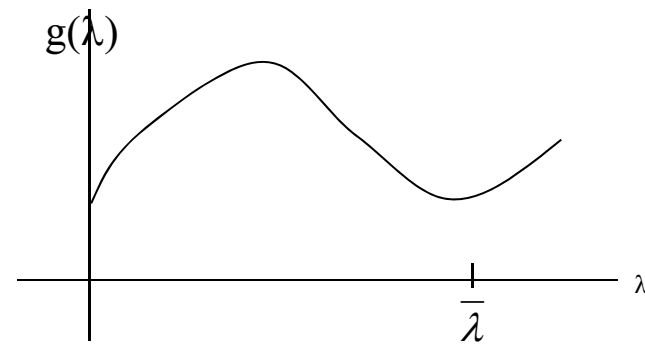
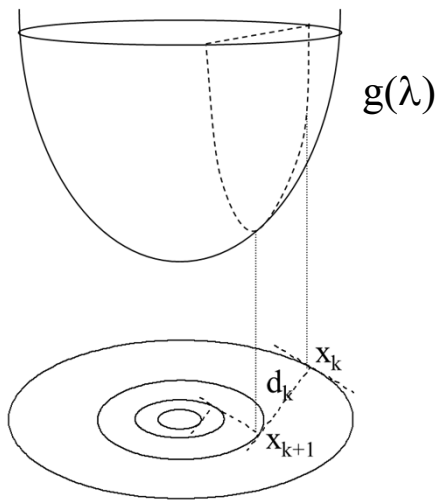
Condiciones necesarias para un mínimo de f en Ω

Sean $x \in \Omega$ y $d \in \mathbb{R}^n$,

d es una **dirección factible** (ó posible) en x si $\exists \bar{\lambda} > 0$ t.q. $x + \alpha d \in \Omega, \forall \alpha, 0 \leq \alpha \leq \bar{\lambda}$.

Idea básica: analizar el comportamiento de f por desplazamientos desde x a lo largo de direcciones factibles.

Dada d factible en $x \in \Omega$, $g(\lambda) = f(x(\lambda))$ es una **función de una única variable**:



Notación y resultados generales

Gradiente de f en x: $\nabla f(x) = \left(\frac{\partial f}{\partial x_1} \quad \dots \quad \frac{\partial f}{\partial x_i} \quad \dots \quad \frac{\partial f}{\partial x_n} \right)$

Hessiano de f en x: $\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_i} & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_i \partial x_1} & \frac{\partial^2 f}{\partial^2 x_i} & \frac{\partial^2 f}{\partial x_i \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_i} & \frac{\partial^2 f}{\partial^2 x_n} \end{pmatrix}$

$A_{n \times n}$ es definida positiva si $d^T A d > 0, \forall d$
(todos los vaps. de $A > 0$)

$A_{n \times n}$ es semi-definida positiva si $d^T A d \geq 0, \forall d$
(todos los vaps. de $A \geq 0$)

Condiciones necesarias para óptimos locales de f en Ω

$\left\{ \begin{array}{l} \text{C. necesarias de 1}^{\text{er}} \text{ orden: } \nabla f(\mathbf{x}). \mathbf{d} \geq 0, \forall \mathbf{d} \text{ factible} \\ \text{C. necesarias de 2}^{\text{o}} \text{ orden: } \mathbf{d}^T \nabla^2 f(\mathbf{x}). \mathbf{d} \geq 0, \forall \mathbf{d} \text{ factible} \end{array} \right.$

Caso no restringido ($\Omega \subseteq \mathbb{R}^n$): \mathbf{d} factible $\forall \mathbf{d}$

C. n. 1 ^{er} orden: $\nabla f(\mathbf{x}) = 0$
C. n. 2 ^o orden: $\nabla^2 f(\mathbf{x})$ semidef. positivo ($\mathbf{d}^T \nabla^2 f(\mathbf{x}). \mathbf{d} \geq 0, \forall \mathbf{d}$)

Ejemplo

$$\text{Min } f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 - 3 x_2$$

$$f(x) = \frac{1}{2} x^T Q x - b x$$

$$Q = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

$$\nabla f(x) = Qx - b$$

$$\nabla^2 f(x) = Q$$

$$\nabla f(x) = 0 \Leftrightarrow Qx - b = 0 \Leftrightarrow \begin{cases} 2x_1 - x_2 = 0 \\ -x_1 + 2x_2 - 3 = 0 \end{cases} \Leftrightarrow \begin{cases} 2x_1 = x_2 \\ -x_1 + 4x_1 = 3 \Rightarrow x_1 = 1, x_2 = 2 \end{cases}$$

(1,2) (es el único punto que cumple las condiciones necesarias de 1er orden)

Es un mínimo global ?

Condición necesaria 2º orden: $d^T \nabla^2 f(x)$. $d \geq 0$, $\forall d$

Hessiano semi-definido positivo

$$Q = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

$$\begin{pmatrix} d_1 & d_2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = (2d_1 - d_2 \quad -d_1 + 2d_2) \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = 2d_1^2 - d_1d_2 - d_1d_2 + 2d_2^2$$

$$|Q - \lambda I| = \begin{vmatrix} 2 - \lambda & -1 \\ -1 & 2 - \lambda \end{vmatrix} = 0 \Leftrightarrow (2 - \lambda)^2 - 1 = 0 \Leftrightarrow \lambda^2 - 4\lambda + 3 = 0 \Rightarrow \begin{cases} \lambda = 3 \\ \lambda = 1 \end{cases}$$

(1,2) (cumple las condiciones necesarias de 2º orden)

Condiciones suficientes 2º orden (caso no restringido)

$$\left\{ \begin{array}{l} \text{i) } \nabla f(x^*) = 0. \\ \text{ii) } \nabla^2 f(x^*) \text{ definido positivo } (d^T \nabla^2 f(x) d > 0, \forall d) \end{array} \right.$$

$\Rightarrow x^*$ es un mínimo relativo estricto de f .

Cuando x^* t.q. $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*) \geq 0$, en general no se puede afirmar nada.

$$\text{a) } f(x,y) = x^4 - y^4; \quad \nabla f(x) = 0 \Leftrightarrow \begin{cases} 4x^3 = 0 \\ -4y^3 = 0 \end{cases} \Leftrightarrow x=(0,0)$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & -12y^2 \end{pmatrix} \quad |\nabla^2 f(x,y)| = -144 x^2 y^2 \Rightarrow \nabla^2 f(x,y) \text{ es indefinido} \\ (\text{valores propios son } 12x^2 > 0; -12y^2 < 0)$$

$$\nabla^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{es semidefinido positivo. Pero } (0,0) \text{ es un punto silla.}$$

$$\text{b) } f(x,y) = x^4 + y^4; \quad \nabla f(x) = 0 \Leftrightarrow \begin{cases} 4x^3 = 0 \\ 4y^3 = 0 \end{cases} \Leftrightarrow x=(0,0)$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{pmatrix}$$

Como antes $\nabla^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ es semidefinido positivo. Ahora $(0,0)$ es un mínimo global.

$$\text{a) } f(x,y) = x^4 - y^4; \quad \nabla f(x) = 0 \Leftrightarrow \begin{cases} 4x^3 = 0 \\ -4y^3 = 0 \end{cases} \Leftrightarrow x=(0,0)$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & -12y^2 \end{pmatrix} \quad |\nabla^2 f(x,y)| = -144 x^2 y^2 \Rightarrow \nabla^2 f(x,y) \text{ es indefinido} \\ (\text{valores propios son } 12x^2 > 0 \text{ y } -12y^2 < 0)$$

$$\nabla^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{es semidefinido positivo. Pero } (0,0) \text{ es un punto silla.}$$

$$\text{b) } f(x,y) = x^4 + y^4; \quad \nabla f(x) = 0 \Leftrightarrow \begin{cases} 4x^3 = 0 \\ 4y^3 = 0 \end{cases} \Leftrightarrow x=(0,0)$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{pmatrix}$$

Como antes $\nabla^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ es semidefinido positivo. Ahora $(0,0)$ es un mínimo global.

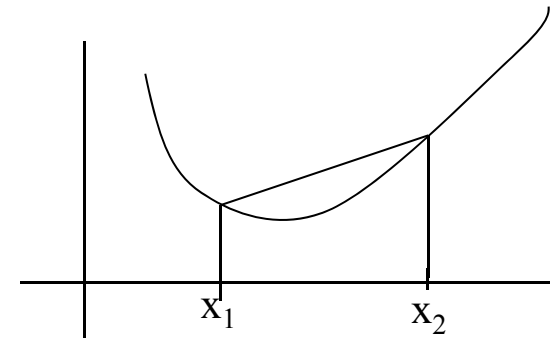
Convexidad

Definición. $f: \Omega \rightarrow \mathbb{R}$, Ω convexo, se dice que es convexa si $\forall x_1, x_2 \in \Omega$,

$$\forall \alpha, 0 \leq \alpha \leq 1, f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$$

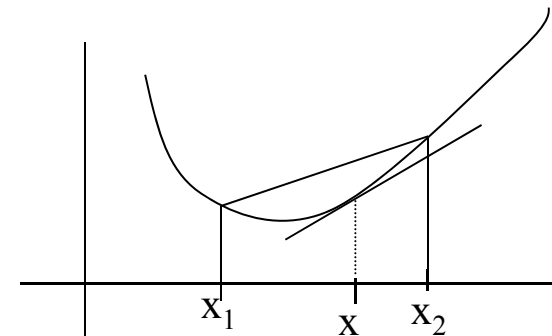
Propiedad: f_1 y f_2 convexas en Ω convexo y $\lambda_1 > 0$, $\lambda_2 > 0$,

$\Rightarrow \lambda_1 f_1 + \lambda_2 f_2$ también es convexa en Ω .



Proposición (Caracterización de funciones convexas): Sea $f \in C^1$ definida en un dominio

Ω convexo. f es convexa $\Leftrightarrow \forall x, y \in \Omega \ f(y) \geq f(x) + \nabla f(x)(y-x)$



Caso particular f convexa y Ω convexo

- f convexa en Ω convexo. Cualquier mínimo relativo de f es un mínimo global.
- $f \in C^1$ convexa en Ω convexo.

Si $\exists x^* \in \Omega$ t.q. $\forall y \in \Omega \nabla f(x^*)(y-x^*) \geq 0 \Rightarrow x^*$ es un mínimo global de f en Ω .

Cuando $f(x)$ es convexa y Ω es convexo,
las condiciones necesarias de 1er. orden son tambien suficientes

Convergencia De Algoritmos

- Un algoritmo es *convergente* si la sucesión de puntos generados por el algoritmo converge en un número finito o infinito de pasos a una solución del problema de optimización que queramos resolver.
- Un algoritmo tiene *convergencia global* si para puntos de partida arbitrarios se converge a una solución.

Convergencia local. Orden de convergencia

Sea $\{r_k\}_{k=1}^{\infty} \longrightarrow r^*$. Se llama *orden de convergencia* de $\{r_k\}_{k=1}^{\infty}$ al mayor número no negativo p tal que

$$\lim_{k \longrightarrow \infty} \frac{|r_{k+1} - r^*|}{|r_k - r^*|^p} < \infty$$

Valores mayores de p indican una convergencia más rápida.
(la distancia al límite r^* , se reduce en cada paso en la p -ésima potencia)

La mayoría de los algoritmos tienen orden de convergencia 1 (a veces 2).

Sea $\{r_k\}_{k=1}^{\infty} \longrightarrow r^*$ tal que $\lim_{k \longrightarrow \infty} \frac{|r_{k+1} - r^*|}{|r_k - r^*|} = \beta < 1$.

Se dice que $\{r_k\}_{k=1}^{\infty}$ *converge linealmente* a r^* con una *tasa de convergencia* β .

Si $\beta=0$, se dice que la convergencia es superlineal.

$$1. \quad \{r_k\}_{k=1}^{\infty} = \{a^k\}_{k=1}^{\infty} \longrightarrow 0, \quad 0 < a < 1 \quad 0 < \lim_{k \longrightarrow \infty} \frac{|a^{k+1} - 0|}{|a^k - 0|^p} = a < \infty$$

p=1 ; orden 1 y tasa a .

$$2. \quad \{r_k\}_{k=1}^{\infty} = \{a^{2^k}\}_{k=1}^{\infty} \longrightarrow 0, \quad 0 < a < 1 \quad 0 < \lim_{k \longrightarrow \infty} \frac{|a^{2^{(k+1)}} - 0|}{|a^{2^k} - 0|^p} = \lim_{k \longrightarrow \infty} \frac{a^{2 \cdot 2^k}}{a^{p \cdot 2^k}} = 1 < \infty$$

p=2. orden 2.

$$3. \quad \{r_k\}_{k=1}^{\infty} = \left\{\frac{1}{k}\right\}_{k=1}^{\infty} \longrightarrow 0, \quad 0 < \lim_{k \longrightarrow \infty} \frac{\left|\frac{1}{k+1} - 0\right|}{\left|\frac{1}{k} - 0\right|^p} = \lim_{k \longrightarrow \infty} \frac{k}{k+1} = 1 < \infty$$

p=1. orden 1 pero no es lineal

$$4. \quad \{r_k\}_{k=1}^{\infty} = \left\{\left(\frac{1}{k}\right)^k\right\}_{k=1}^{\infty} \longrightarrow 0,$$

p=1; Es superlineal

$$0 < \lim_{k \longrightarrow \infty} \frac{\left|\left(\frac{1}{k+1}\right)^{k+1} - 0\right|}{\left|\left(\frac{1}{k}\right)^k - 0\right|^p} = \lim_{k \longrightarrow \infty} \frac{(k)^k}{(k+1)^{(k+1)p}} = \lim_{k \longrightarrow \infty} \left(\frac{k}{k+1}\right)^k \frac{1}{k+1} = 0$$

Exploración lineal

Procedimientos numéricos para minimizar funciones unidimensionales
(proceso para determinar el mínimo de una función a lo largo de una dirección)

Ejemplo: $f(x,y)=x^2y$. $x_0=(2,5)$, $d=(d_1, d_2)$

$$x(\lambda)=(2,5) + \lambda (d_1, d_2) = (2- \lambda d_1, 5- \lambda d_2) ; g(\lambda)=(2- \lambda d_1)^2(5- \lambda d_2)$$

En general $g(\lambda)$ no puede minimizarse de forma analítica.

Min $g(\lambda)$ se realiza buscando de “forma inteligente” a lo largo de la línea.

(embudo de algoritmos de optimización no lineal puesto que, una exploración lineal debe realizarse en cada iteración).

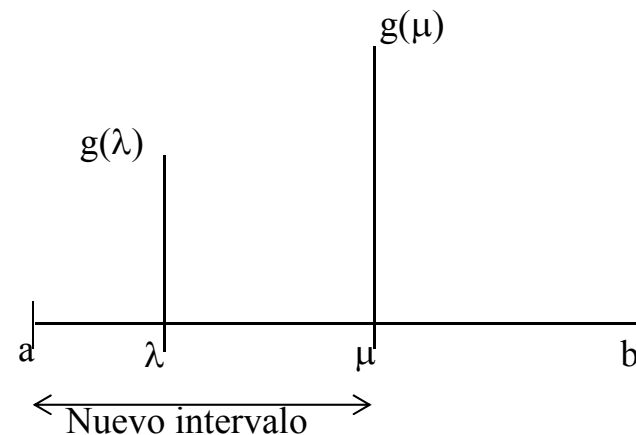
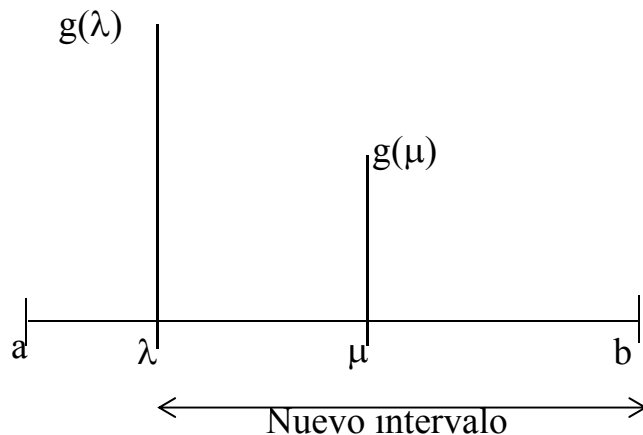
Un *intervalo de incertidumbre* para $g : R \longrightarrow R$, es un intervalo $[a, b]$ que contiene al $\min g(\lambda)$. (Contiene 3 puntos en U y g es *unimodal* en $[a, b]$)

Métodos de exploración lineal: método iterativo que en cada paso reduce la amplitud del intervalo de incertidumbre, eliminando una porción del intervalo de las que se sabe que no contiene al mínimo.

Sea $[a, b]$ un intervalo de incertidumbre para $g : R \longrightarrow R$. Sean $\lambda, \mu \in [a, b]$ tales que $\lambda < \mu$.

Si $g(\lambda) > g(\mu) \Rightarrow g(\gamma) \geq g(\mu) \quad \forall \gamma \in [a, \lambda)$

Si $g(\lambda) \leq g(\mu) \Rightarrow g(\gamma) \geq g(\lambda) \quad \forall \gamma \in (\mu, b]$



Método de Fibonacci

Obtendremos el mínimo de g en $[a, b]$ evaluando g en ciertos puntos.

Seleccionar N puntos para evaluar g de forma que el intervalo de incertidumbre se reduzca lo más posible al final de las N evaluaciones.

Estrategia óptima:

$d_1 = b - a$ la amplitud del intervalo de incertidumbre inicial

d_k : amplitud del intervalo de incertidumbre después de k evaluaciones

Propiedad: $d_k = \frac{F_{N-k+1}}{F_N} d_1$, F_k : números de Fibonacci ($F_{k+1} = F_k + F_{k-1}$, $F_0 = F_1 = 1$)
(1, 1, 2, 3, 5, 8, 13, 21, 34, ...)

Primera iteración:

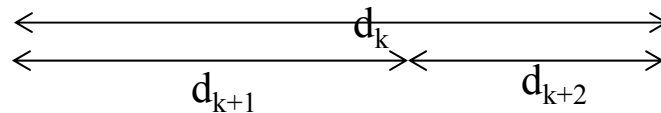
- Dos evaluaciones de g en los dos puntos de $[a, b]$ cuyas distancias a a y b es d_1 .
- Se conoce el valor de g en 4 puntos.
- Se reduce el intervalo $[a, b]$ según el resultado de la transparencia anterior.

Iteraciones posteriores:

- Se conoce el valor de g en 3 puntos (los dos extremos y uno en el interior).
- Se realiza una única evaluación adicional, en el punto a distancia d_k del extremo del intervalo que se ha actualizado en la iteración anterior.
- Después de realizar la evaluación de g , se vuelve a aplicar resultado transparencia anterior para actualizar el correspondiente extremo del intervalo.

Método de Fibonacci: Propiedades

- Se conoce a priori la amplitud del intervalo de incertidumbre después de la k -ésima evaluación de g .
- En cada iteración se cumple la relación: $d_k = d_{k+1} + d_{k+2}$,



- En cada iteración los puntos $\lambda, \mu \in [a, b]$ en los que se conoce el valor de g están simétricos respecto a a y b . Es decir: $\lambda - a = b - \mu$.
- En cada iteración (excepto la primera) se realiza una única evaluación de g .
- Es un método óptimo, en el sentido de que minimiza amplitud del intervalo de incertidumbre después de k evaluaciones
- En la última iteración hay que evaluar la función en un punto en el que ya se ha evaluado previamente. Es la llamada *irregularidad de Fibonacci*. Lo que se hace es evaluar la función en un punto que esté $\pm \varepsilon$ del punto que toca.

Ejemplo

$$f(x,y)=x^2y. \quad x_0=(2,5), \quad d=(-4, -4)$$

$$x(\lambda)=(2,5) + \lambda (-4, -4) = (2-4\lambda, 5-4\lambda) ; \quad g(\lambda)=(2-4\lambda)^2(5-4\lambda)$$

$$[0, 1] \text{ es un intervalo de incertidumbre. } g(0)=20, \quad g(1)=4. \quad d_1 = \sqrt{16+16} = 4\sqrt{2}$$

N=5 evaluaciones de Fibonacci

$$d_2 = \frac{F_4}{F_5} d_1 = \frac{5}{8} d_1, \quad d_3 = \frac{3}{8} d_1, \quad d_4 = \frac{2}{8} d_1, \quad d_5 = \frac{1}{8} d_1.$$

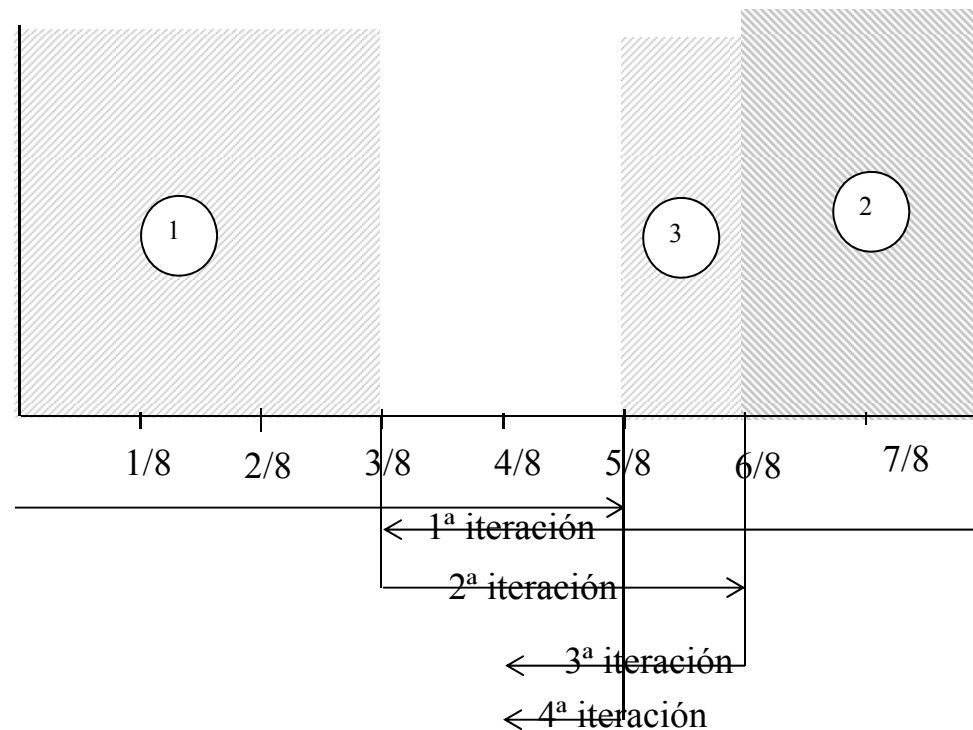
$$g(3/8)=7/8$$

$$g(5/8)=5/8$$

$$g(3/8+3/8) = g(6/8)=2$$

$$g(6/8-2/8) = g(4/8)=0$$

$$g(5/8-1/8^*) = g(1/2-\varepsilon) > 0$$



Métodos basados en las derivadas

Si g es diferenciable se pueden tener métodos de exploración lineal con más información que la basada en las evaluaciones de g .

Ajustes por curvas: Realizar sucesivas aproximaciones de g mediante polinomios de distintos grados cuyo mínimo es fácil de calcular.

Ajustes Cuadráticos. Aproximamos $g(\lambda)$ mediante $q(\lambda) = a \lambda^2 + b \lambda + c$.

El mínimo de $q(\lambda)$ es $q'(\lambda) = 0 \Leftrightarrow 2 a \lambda + b = 0 \Leftrightarrow \lambda = -\frac{b}{2a}$
 $q''(\lambda) = 2a > 0 \Leftrightarrow a > 0$.

. Suponemos conocido el valor de g en 2 puntos y la derivada en uno de ellos.

$$g(\lambda_1) = q(\lambda_1) = a \lambda_1^2 + b \lambda_1 + c$$

$$g(\lambda_2) = q(\lambda_2) = a \lambda_2^2 + b \lambda_2 + c$$

$$g'(\lambda_1) = \nabla f(x(\lambda_1)).d = q'(\lambda_1) = 2 a \lambda_1 + b$$

Sistema de 3 ecuaciones con 3 incógnitas a , b y c .

. Suponemos conocido el valor de g en 3 puntos.

$$g(\lambda_1) = q(\lambda_1) = a \lambda_1^2 + b \lambda_1 + c$$

$$g(\lambda_2) = q(\lambda_2) = a \lambda_2^2 + b \lambda_2 + c$$

$$g(\lambda_3) = q(\lambda_3) = a \lambda_3^2 + b \lambda_3 + c$$

Sistema de 3 ecuaciones con 3 incógnitas a , b y c .

Ajustes Cúbicos. Se aproxima g mediante $q(\lambda) = a \lambda^3 + b \lambda^2 + c \lambda + d$

El mínimo de $q(\lambda)$ es $q'(\lambda) = 0 \Leftrightarrow 3 a \lambda^2 + 2 b \lambda + c = 0 \Leftrightarrow$

$$\lambda = \frac{-2b \pm \sqrt{4b^2 - 12ac}}{6a} = \frac{-2 \pm \sqrt{b^2 - 3ac}}{3a}$$

El λ que da el mínimo es $q''(\lambda) > 0 \Leftrightarrow 6 a \lambda + 2 b > 0$, es $\lambda = \frac{-2 + \sqrt{b^2 - 3ac}}{3a}$

Debe existir $\sqrt{b^2 - 3ac} \Leftrightarrow b^2 - 3ac > 0$

. Suponemos conocido el valor de g en 3 puntos y la derivada en uno de ellos.

$$g(\lambda_1) = q(\lambda_1) = a \lambda_1^3 + b \lambda_1^2 + c \lambda_1 + d$$

$$g(\lambda_2) = q(\lambda_2) = a \lambda_2^3 + b \lambda_2^2 + c \lambda_2 + d$$

$$g(\lambda_3) = q(\lambda_3) = a \lambda_3^3 + b \lambda_3^2 + c \lambda_3 + d$$

$$g'(\lambda_1) = \nabla f(x(\lambda_1)).d = q'(\lambda_1) = 3 a \lambda_1^2 + 2 b \lambda_1 + c$$

Sistema de 4 ecuaciones con 4 incógnitas a, b, c y d .

. Suponemos conocido el valor de g en 2 puntos y la derivada en los dos puntos.

$$g(\lambda_1) = q(\lambda_1) = a \lambda_1^3 + b \lambda_1^2 + c \lambda_1 + d$$

$$g(\lambda_2) = q(\lambda_2) = a \lambda_2^3 + b \lambda_2^2 + c \lambda_2 + d$$

$$g'(\lambda_1) = \nabla f(x(\lambda_1)).d = q'(\lambda_1) = 3 a \lambda_1^2 + 2 b \lambda_1 + c$$

$$g'(\lambda_2) = \nabla f(x(\lambda_2)).d = q'(\lambda_2) = 3 a \lambda_2^2 + 2 b \lambda_2 + c$$

Sistema de 4 ecuaciones con 4 incógnitas a, b, c y d .

Exploraciones aproximadas.

Es importante minimizar $g(\lambda) = f(x(\lambda)) = f(x^a + \lambda d)$ de la forma más eficiente posible. Aunque en cada iteración se obtenga el mínimo $g(\lambda)$ aproximadamente, se puede mantener la convergencia global de los métodos que la tengan.

Veamos 2 ejemplos para justificar la necesidad de las condiciones que se exigirán.

Ejemplo 1

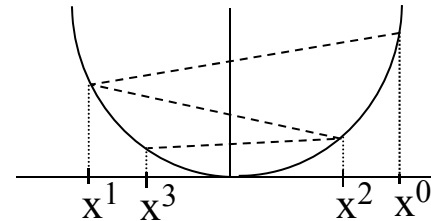
$$f(x)=x^2;$$

$$x^0=2, d_k=(-1)^{k+1}, \lambda_k=2 + 3 (2^{-(k+1)}).$$

$$x^k=\{2, -3/2, 5/4, -9/8, \dots\} = \{(-1)^k(1+2^{-k})\}$$

$f(x^k)=(-1)^{2k}(1+2^{-k})^2 = (1+2^{-k})^2$ es estrictamente monótona decreciente y convergente.

Pero ... $\lim_{k \rightarrow \infty} f(x^k) = 1 \neq 0 = f(0)$



No es suficiente con exigir un decremento estricto del valor de la función en cada iteración.

En este ejemplo el problema es que las longitudes de paso son excesivamente grandes respecto al decremento de la función que se obtiene.

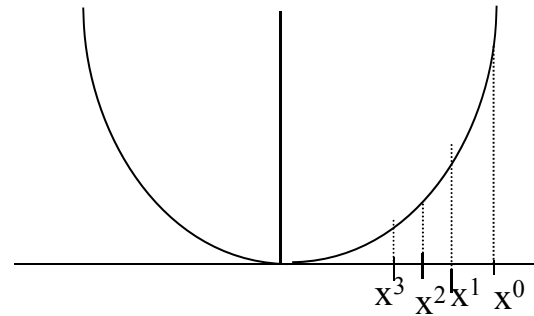
Ejemplo 2

$$f(x)=x^2; x^0=2, d_k=-1, \lambda_k=2^{-k+1}. x^k=\{2, 3/2, 5/4, 9/8, \dots\} = \{1+2^{-k}\}$$

$f(x_k) = (1+2^{-k})^2$ es estrictamente monótona decreciente y convergente.

Pero ...

$$\lim_{k \rightarrow \infty} f(x^k) = 1 \neq 0 = f(0)$$

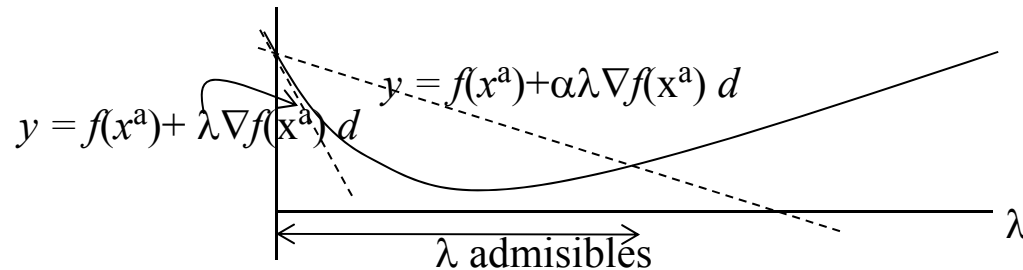


En este ejemplo el problema es que las longitudes de paso son excesivamente cortas.

Condiciones de Armijo y Goldstein I

Para evitar pasos excesivamente largos con decrementos proporcionalmente demasiado pequeños se impone que el decremento medio de la función entre x^a y x^+ sea por lo menos una fracción prefijada del decremento inicial de la función en la dirección dada. Es decir, tomamos $\alpha \in (0,1)$ y elegimos $\lambda > 0$ t.q.

$$f(x^+) \leq f(x^a) + \alpha \lambda \nabla f(x^a) d.$$



Ejemplo

En el primer ejemplo, tomando $x^3 = -9/8$ y $x^4 = 17/16$ vemos que en el tercer paso se cumple la primera condición de Armijo-Goldstein para $\alpha = 0.1$.

$$f(x^3) = (1+2^{-3})^2 ; f(x^4) = (1+2^{-4})^2$$

$$y = f(x^3) + \alpha \lambda_3 \nabla f(x^3) d_3 = (1+2^{-3})^2 + \alpha [2 + 3(2^{-4})](2(-9/8))(-1)^4 =$$

$$= \left(1 + \frac{1}{2^3}\right)^2 - \alpha \frac{9}{4} \left[2 + \frac{3}{2^4}\right]$$

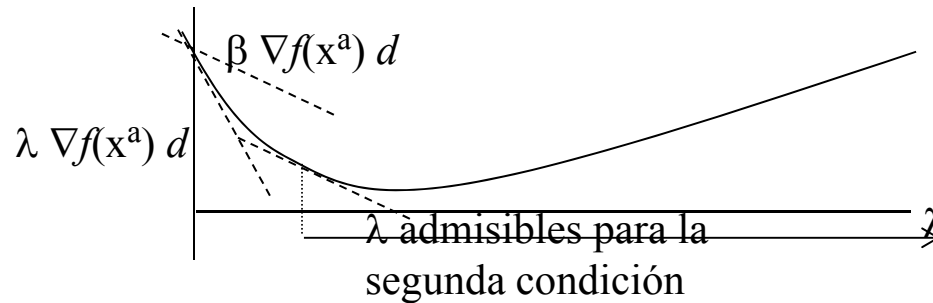
$$\left(1 + \frac{1}{2^4}\right)^2 \geq \left(1 + \frac{1}{2^3}\right)^2 - \alpha \frac{9}{4} \left[2 + \frac{3}{2^4}\right] \Leftrightarrow \frac{0.9}{4} \left[2 + \frac{3}{2^4}\right] \geq \left(2 + \frac{1}{2^3} + \frac{1}{2^4}\right) \left(\frac{1}{2^3} - \frac{1}{2^4}\right) \Leftrightarrow \frac{0.9}{4} \geq \left(\frac{1}{8} - \frac{1}{16}\right) = \frac{1}{16}$$

$$f(x^+) \leq f(x^a) + \alpha \lambda \nabla f(x^a) d.$$

Condiciones de Armijo y Goldstein II

Para evitar pasos excesivamente pequeños exigiremos que el decremento de la función en x^+ sea mayor o igual que una fracción prefijada del decremento de la función en x^a . Es decir para alguna constante prefijada $\beta \in (\alpha, 1)$.

$$\nabla f(x^+).d = \nabla f(x^a + \lambda d) \nabla \geq \beta \nabla f(x^a) d$$



$\beta > \alpha$ garantiza que las dos condiciones pueden cumplirse simultáneamente.

Métodos generales de optimización

Método del gradiente

En cada iteración se selecciona como dirección de movimiento

$$d_k = - \nabla f(x^k)^T.$$

La sucesión de puntos generada es: $x^{k+1} = x_k - \lambda_k (\nabla f(x^k))^T$

$$\text{donde } \lambda_k \text{ t.q. } f(x_k + \lambda_k d_k) = \text{Min}_{\lambda} f(x^k + \lambda d_k)$$

➤ Es un algoritmo de descenso: $f(x_{k+1}) < f(x_k)$

➤ Convergencia global: $\lim_{k \rightarrow \infty} x^k = x^*$ donde x^* es mínimo relativo de $f(x)$.

(para mantener la convergencia global basta con que λ_k cumpla las condiciones de Armijo y Goldstein)

➤ Convergencia local: $\{f(x^k)\} \rightarrow f(x^*)$ linealmente con ratio no mayor que $\left(\frac{A-a}{A+a} \right)^2$

donde $A > 0$ y $a > 0$ los vaps mayor y menor respectivamente de $\nabla^2 f(x^*)$

(es necesario que $f \in C^2$ con un mínimo relativo en x^*)

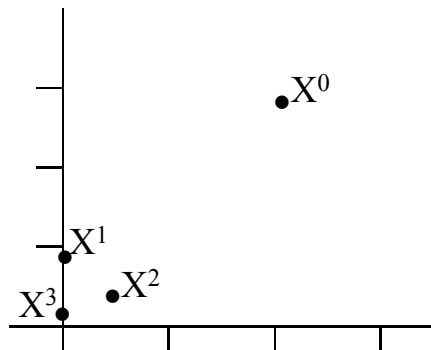
Ejemplo

$$f(x,y)=2(x^2+(y-x)^2) \quad Q=\begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix} \quad \nabla f(x,y)=(4(2x-y), 4(-x+y)) \quad X_0=\begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad f(X^0)=10$$

$$\nabla f(X^0) = (4, 4) \quad \lambda_0 = \frac{\|(4 \ 4)\|^2}{(4 \ 4) \begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix} \begin{pmatrix} 4 \\ 4 \end{pmatrix}} = \frac{1}{2} \quad X^1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad f(X^1)=2$$

$$\nabla f(X^1) = (-4, 4) \quad \lambda_1 = \frac{\|(-4 \ 4)\|^2}{(-4 \ 4) \begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix} \begin{pmatrix} -4 \\ 4 \end{pmatrix}} = \frac{1}{10} \quad X^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \frac{1}{10} \begin{pmatrix} -4 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{4}{10} \\ \frac{6}{10} \end{pmatrix} \quad f(X^2)=2/5$$

$$\nabla f(X^2) = (4/5, 4/5) \quad \lambda_1 = \frac{\|(4/5 \ 4/5)\|^2}{(4/5 \ 4/5) \begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix} \begin{pmatrix} 4/5 \\ 4/5 \end{pmatrix}} = \frac{1}{2} \quad X^3 = \begin{pmatrix} \frac{4}{10} \\ \frac{6}{10} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 4/5 \\ 4/5 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{2}{10} \end{pmatrix} \quad f(X^3)=2/25$$



Método de Newton.

En cada iteración se selecciona como dirección de movimiento

$$d_k = -[\nabla^2 f(X^k)]^{-1} (\nabla f(X^k))^T.$$

$$X^{k+1} = X^k - [\nabla^2 f(X^k)]^{-1} (\nabla f(X^k))^T; \quad \lambda_k = 1$$

En cada iteración se realiza una aproximación cuadrática, $q(x)$, de f en x^k .

x_{k+1} : óptimo de la aproximación cuadrática.

$$\left. \begin{array}{l} q(X^k) = f(X^k); \\ \nabla q(X^k) = \nabla f(X^k); \\ \nabla^2 q(X^k) = \nabla^2 f(X^k) \end{array} \right\} \longrightarrow q(X) = f(X^k) + \nabla f(X^k)^T (X - X^k) + \frac{1}{2} (X - X^k)^T \nabla^2 f(X^k) (X - X^k).$$

El mínimo de $q(x)$ es:

$$\nabla q(X) = 0 \Leftrightarrow (\nabla f(X^k))^T = \nabla^2 f(X^k) (X - X^k) \Leftrightarrow$$

$$X^{k+1} = X^k - [\nabla^2 f(X^k)]^{-1} (\nabla f(X^k))^T$$

Propiedades del Método de Newton

- No tiene convergencia global.
Si $[\nabla^2 f(X^k)]^{-1}$ no definido positivo, d_k puede no ser dirección de descenso ($f(X^{k+1})$ puede ser mayor que $f(X^k)$).
- Convergencia Local: Si x^0 está *suficientemente próximo* a x^* y $\nabla^2 f(x^*)$ definido positivo, la sucesión de puntos generada por el método de Newton converge a x^* con un orden de convergencia de por lo menos 2. ($f \in C^3$ con mínimo relativo en x^*).
- En el caso de funciones cuadráticas en un paso se obtiene el óptimo
- Cuando no es definido positivo, puede generar un punto peor.
- Para garantizar que la dirección sea de descenso es necesario perturbar $[\nabla^2 f(X^k)]^{-1}$.
- Para garantizar la convergencia global, podemos introducir el parámetro de longitud de paso y realizar exploración lineal (en los puntos próximos a x^* , $\lambda_k \approx 1$)

Ejemplo

$$f(x, y) = x^4 + 2x^2y^2 + y^4. \quad \nabla f(x, y) = (4x^3 + 4xy^2, 4y^3 + 4x^2y);$$

$$\nabla^2 f(x, y) = \begin{pmatrix} 12x^2 + 4y^2 & 8xy \\ 8xy & 4x^2 + 12y^2 \end{pmatrix} \quad X^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad ; \quad \nabla f(X^0) = (8, 8); \quad H_0 = \begin{pmatrix} 16 & 8 \\ 8 & 16 \end{pmatrix}$$

$$\text{Resolvemos el sistema } -\nabla f(X^0)^T = H_0 d \Leftrightarrow -\begin{pmatrix} 8 \\ 8 \end{pmatrix} = \begin{pmatrix} 16 & 8 \\ 8 & 16 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$$

$$\text{Calculamos la descomposición TDT}^T \text{ de } H_0. \quad \begin{pmatrix} 16 & 8 \\ 8 & 16 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 16 & \\ & 12 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{pmatrix}$$

$$-\begin{pmatrix} 8 \\ 8 \end{pmatrix} = \begin{pmatrix} 16 & 8 \\ 8 & 16 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \Leftrightarrow -\begin{pmatrix} 8 \\ 8 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 16 & \\ & 12 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 16 & \\ 8 & 12 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

$$\text{Primero resolvemos por sustitución inversa el sistema } \begin{cases} -8 = 16h_1 \\ -8 = 8h_1 + 12h_2 \end{cases} \Leftrightarrow \begin{cases} h_1 = -\frac{1}{2} \\ h_2 = -\frac{1}{3} \end{cases}$$

$$\text{Después resolvemos (también por sustitución inversa)} \quad \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{3} \end{pmatrix} \Leftrightarrow \begin{cases} d_1 + \frac{1}{2}d_2 = -\frac{1}{2} \\ d_2 = -\frac{1}{3} \end{cases} \Leftrightarrow \begin{cases} d_1 = -\frac{1}{3} \\ d_2 = -\frac{1}{3} \end{cases}$$

$$\Rightarrow X^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}$$

Modificaciones del Método de Newton para que la dirección sea de descenso

Tomar $d_k = -[\nabla^2 f(X^k) + \varepsilon_n I_n]^{-1} (\nabla f(X^k))^T$ tal que

$[\nabla^2 f(X^k) + \varepsilon_n I_n]^{-1}$ “suficientemente” definido positivo.

Para ello,

- Fijar $\delta > 0$,
- Calcular los valores propios de $H_k = \nabla^2 f(X^k)$
- Tomar ε_n como la menor constante t.q. los vap's de $\nabla^2 f(X^k) + \varepsilon_n I_n$ sean $\geq \delta$.

Propiedades:

- Tiene convergencia global
- Si $\delta > 0 < \text{“menor valor propio de } H_k\text{”}$, para puntos X_k suficientemente próximos a x^* , ε_n será cero y la convergencia local es de orden 2.
- ¿Cómo elegir δ ?
 - Si δ pequeño, podemos tener matrices casi singulares
 - Si δ grande, podemos perder la convergencia local de orden 2.
- Inconveniente: Tenemos que calcular los valores propios de H_k . Además, la matriz resultante puede no parecerse en nada a H_k .

Supongamos que $\nabla f(X^k) = (2, 5)$; y que $H_0 = \begin{pmatrix} 1 & 0 \\ 0 & -5 \end{pmatrix}$

La dirección de Newton es $d^T = (-2, 1)$

$$\hat{H}_0 = \begin{pmatrix} 1 + (5 + \delta) & 0 \\ 0 & \delta \end{pmatrix} \Leftrightarrow \begin{pmatrix} 6 + \delta & 0 \\ 0 & \delta \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = - \begin{pmatrix} 2 \\ 5 \end{pmatrix} \Leftrightarrow \begin{cases} d_1 = -2 / (6 + \delta) \\ d_2 = -5 / \delta \end{cases}$$

En la práctica tendremos la descomposición TDT^T de H_k .

Aproximaremos $H_k = TDT^T$ por $\hat{H}_k = T(D + \varepsilon_n I_n)T^T$, donde ε_n es la menor constante t.q. los elementos de $D + \varepsilon_n I_n$ sean $\geq \delta$.

Ejemplo :

$$\begin{pmatrix} 1 & 1 & 2 \\ 1 & 1+10^{-20} & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ & 1 & \\ 2 & 10^{20} & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & 10^{-20} & \\ & & -3-10^{-20} \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ & 1 & 10^{20} \\ & & 1 \end{pmatrix}$$

vap's: 5.1131, 0.0888, -2.2019. Tomando $\delta = 0.05$, $\varepsilon = 0.05 + (3+10^{-20})$, queda

$$\bar{H} = \begin{pmatrix} 1 & & \\ 1 & 1 & \\ 2 & 10^{20} & 1 \end{pmatrix} \left[\begin{pmatrix} 1 & & \\ & 10^{-20} & \\ & & -3-10^{-20} \end{pmatrix} + (3.05+10^{-20})I_3 \right] \begin{pmatrix} 1 & 1 & 2 \\ & 1 & 10^{20} \\ & & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 4.05+10^{-20} & 4.05+10^{-20} & 8.1+2x10^{-20} \\ 4.05+10^{-20} & 7.1+2x10^{-20} & 9.1+2x10^{-20}+3.05x10^{20} \\ 8.1+2x10^{-20} & 9.1+2x10^{-20}+3.05x10^{20} & 10^{20}+4x10^{-20} \end{pmatrix}$$

De nuevo, puede ocurrir que la aproximación H_k sea muy diferente de H_k .

Posibilidad (a veces perturba menos H_k): modificar sólo los elementos de D que sean < 0

Ejemplo: $\nabla f(X^k) = (2, 5)$; $H_0 = \begin{pmatrix} 1 & 0 \\ 0 & -5 \end{pmatrix}$ La dirección de Newton es $d^T = (-2, 1)$.

Modificaremos únicamente el elemento (2,2) de D , $H_0 = (TDT^T)$. Posibilidades:

- Hacer

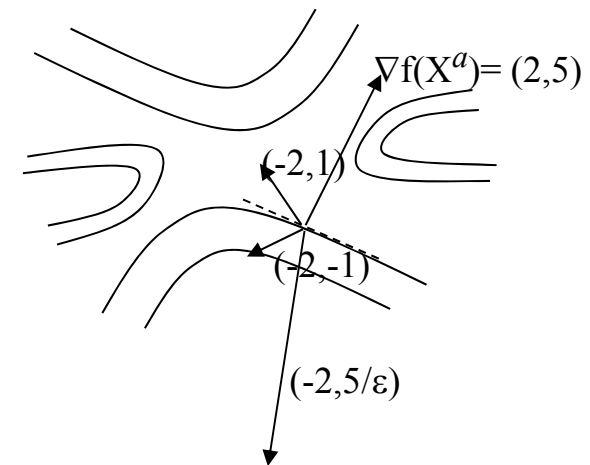
$$\hat{H}_0 = \begin{pmatrix} 1 & 0 \\ 0 & \delta \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & \delta \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = - \begin{pmatrix} 2 \\ 5 \end{pmatrix} \Leftrightarrow \begin{cases} d_1 = -2 \\ d_2 = -5/\delta \end{cases}$$

(muy grande en valor absoluto)

- Hacer

$$\hat{H}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = - \begin{pmatrix} 2 \\ 5 \end{pmatrix} \Leftrightarrow \begin{cases} d_1 = -2 \\ d_2 = -1 \end{cases}$$

(tiene la misma norma que la
dirección original de Newton)



Óptimos locales

$$\left\{ \begin{array}{l} \text{C. necesarias de 1}^{\text{er}} \text{ orden: } \nabla f(x) \cdot d \geq 0, \forall d \text{ factible} \\ \text{C. necesarias de 2}^{\text{o}} \text{ orden: } d^T \nabla^2 f(x) \cdot d \geq 0, \forall d \text{ factible} \end{array} \right.$$

Caso no restringido: d factible $\forall d$

- C. n. 1^{er} orden: $\nabla f(x) = 0$
- C. n. 2^o orden: $\nabla^2 f(x)$ semidef. positivo

Caso restringido: Kuhn-Tucker: $L(x, \lambda, \mu) = f(x) + \lambda g(x) + \mu h(x)$

$$\nabla L(x, \lambda, \mu) = \nabla f(x) + \lambda \cdot \nabla g(x) + \mu \cdot \nabla h(x) = 0$$

$$\mu h(x) = 0$$

$$\mu \geq 0$$

Min $f(x)$

$$g(x) = 0$$

$$h(x) \leq 0$$

$\nabla^2 L(x, \lambda, \mu)$ semidef. positivo en el plano tangente
a las restricciones activas en x

Cuando $f(x)$ es convexa y Ω es convexo,
las condiciones necesarias son tambien suficientes