Aplicaciones de la Optimización Convexa al análisis de redes

Bibliografía optimización convexa:

Nonlinear Programming: 2nd Edition. by Dimitri P. **Bertsekas**. ISBN: 1-886529-00-0. Publication: 1999

Convex Optimization, Stephen P. Boyd, Lieven Vandenberghe

Cambridge University Press, Mar 8, 2004

Introducción

Repaso de conceptos básicos de funciones de varias variables y convexidad

Repaso : Función derivada parcial

- La derivada parcial de f con respecto de x, la denotamos por f_x, y se define así
- para todos los puntos (x, y) donde este límite exista.

$$f_x(x, y) = \lim_{h \to 0} \frac{f(x+h, y)-f(x, y)}{h}$$

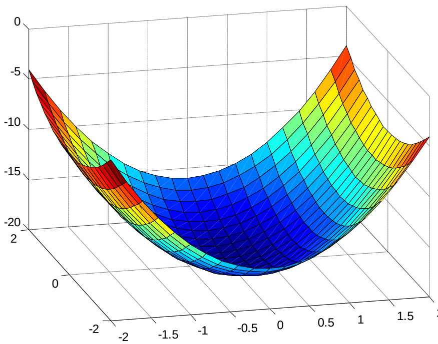
- La derivada parcial de f con respecto de y se define de forma similar
- Si z = f(x, y), otras notaciones usuales para las derivadas parciales son

$$\frac{\partial f}{\partial x}(x, y) = \frac{\partial z}{\partial x} \qquad \frac{\partial f}{\partial y}(x, y) = \frac{\partial z}{\partial y}$$

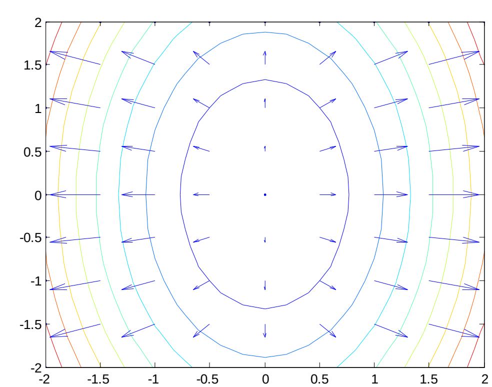
Repaso: Gradiente

- El gradiente de una función escalar de n variables $f(x_1, x_2, \ldots, x_n)$, denotado por ∇f, es el vector n-dimensional

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n} \right)$$



El gradiente de una función en un punto indica la dirección, a partir de ese punto, en la que dicha función crece más rápidamente y, además, la dirección ortogonal a las curvas de nivel de f (curvas en las que la función tiene un valor constante)



Repaso: Derivada direccional

- La Derivada direccional de f en p según el vector unitario μ [$\mathrm{D}_\mu f(p)$] es el producto escalar del gradiente en p, por μ :

$$\mathrm{D}_\mu \; f(p) = \nabla f(p)^{\mathrm{T}} \mu$$

¿En qué sentido deberían desplazarse las variables de f, partiendo del punto p, para que los valores de f crezcan más rápidamente?

Repaso: Derivada direccional

- Como la rapidez está dada por : $\nabla f(p)^{T}\, \mu$

- En esta expresión se suponen ya conocidos f y p; faltando conocer "μ" que haga máximo el producto escalar

- Siendo $\nabla f(p)^{T}\, \mu = \left|\nabla f(p)\right| . \left|\mu\right|$ Cos $\theta = \left|\nabla f(p)\right|$.(1). Cos θ

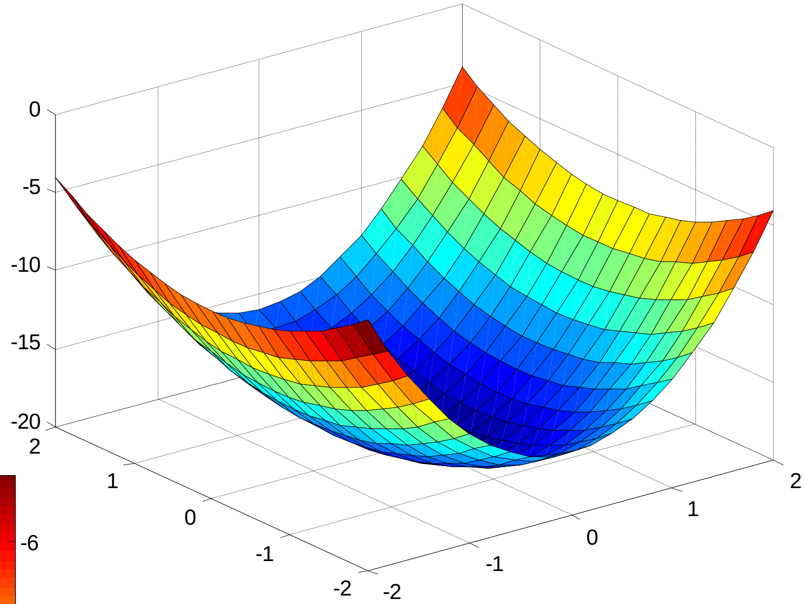
- Donde : θ , es el ángulo formado por los vectores $\nabla f(p)$ y μ

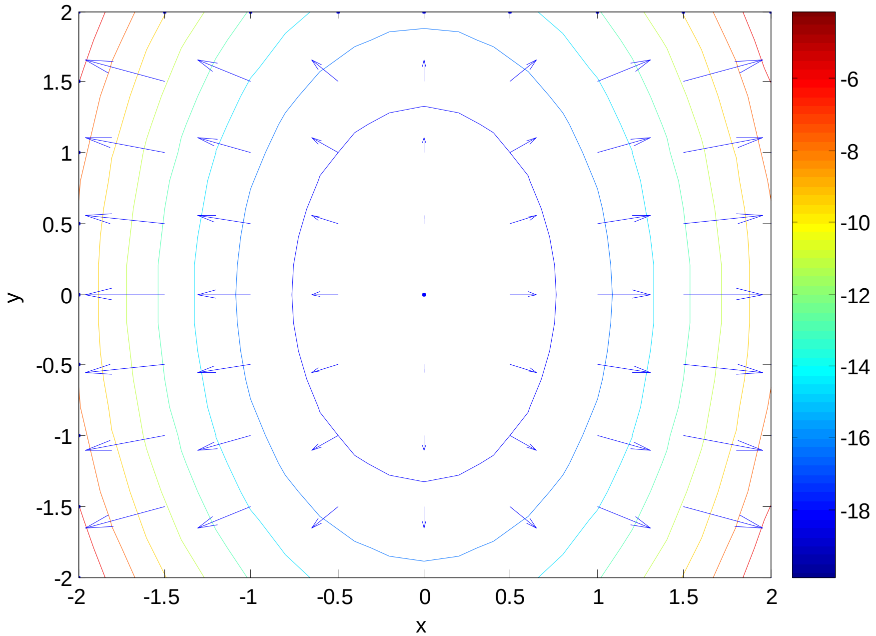
- $\nabla f(p)^{T}\, \mu$, será máximo si y sólo si Cos θ es máximo, ó sea cuando $\theta = 0$ y $\nabla f(p)$ con μ son colineales. Lo cual significa que el vector unitario μ debe tener el mismo sentido que el vector gradiente de f en p

Ejemplo:

$f(x,y) = -20 + 3x^2 + y^2$

$\nabla f = [6x, 2y]$





Repaso: Hessiano

- El Hessiano de una función escalar de n variables $f(x_1,x_2,\ldots, x_n,)$, denotado por H_f, es la matriz de dimensión nxn

$$\boldsymbol{H}_{\boldsymbol{f}} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- Una matriz cuadrada A es *definida positiva* si : $x^T A x > 0 \quad \forall x \neq 0$ y es definida negativa si la desigualdad es la contraria

Repaso: Diferenciabilidad de funciones de varias variables

- ***Definición:***
$$f\left(\vec{x}_0+\vec{v}\right)=f\left(\vec{x}_0\right)+\nabla f^T\left(\vec{x}_0\right)\cdot\vec{v}+\left|\vec{v}\right|E_c(\vec{v})$$
$$E_c(\vec{v})\to 0 \text{ cuando } \vec{v}\to 0$$
- ***Teorema***: Si $f : u$ → R, $u \subset R^n$, es diferenciable en x entonces es continua en x. El recíproco es falso.

- ***Teorema***: (Condición suficiente de diferenciabilidad)
 Si $f : u$ → R, $u \subset R^n$, posee derivadas parciales continuas en $x_0 \in u$ entonces f es diferenciable en x_0. Sin embargo, una función puede ser diferenciable en un punto sin que sus derivadas parciales sean continuas en dicho punto.

- ***Definición***: Decimos que una función es de clase C^k en $A \subseteq R^n$, y escribimos $f \in C^k(A)$, si todas sus derivadas parciales de orden k existen y son continuas en A. *La llamaremos también k veces continuamente diferenciable. Esta nomenclatura no es uniforme en diferentes autores.*

Repaso: Formula de Taylor en Varias Variables

$$f(\vec{x}) = f(\vec{x}_0) + \sum_{j=1}^{n} \frac{\partial f(x_0)}{\partial x_j}(x_j - x_{oj}) + \frac{1}{2!}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_{0i})\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}(x_j - x_{0j}) + \cdots$$

- En una notación mas compacta

$$f(\vec{x}) = f(\vec{x}_0) + \nabla f^T(\vec{x}_0)\cdot(\vec{x} - \vec{x}_0) + \frac{1}{2!}(\vec{x} - x_0)^T \overline{\overline{H}}(\vec{x} - \bar{x}_0) + \ldots$$

donde,

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right)$$

Gradiente

$$H \equiv \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{vmatrix}$$

Hessiano

Optimización

Introducción

Optimización

Formulación del problema de optimización

$$\min_{x\in X} f(x)$$

$f:\Re^n \rightarrow \Re$ es una función continua de n variables

X= $\Re^n$ *o* X es un subconjunto de $\Re^n$.

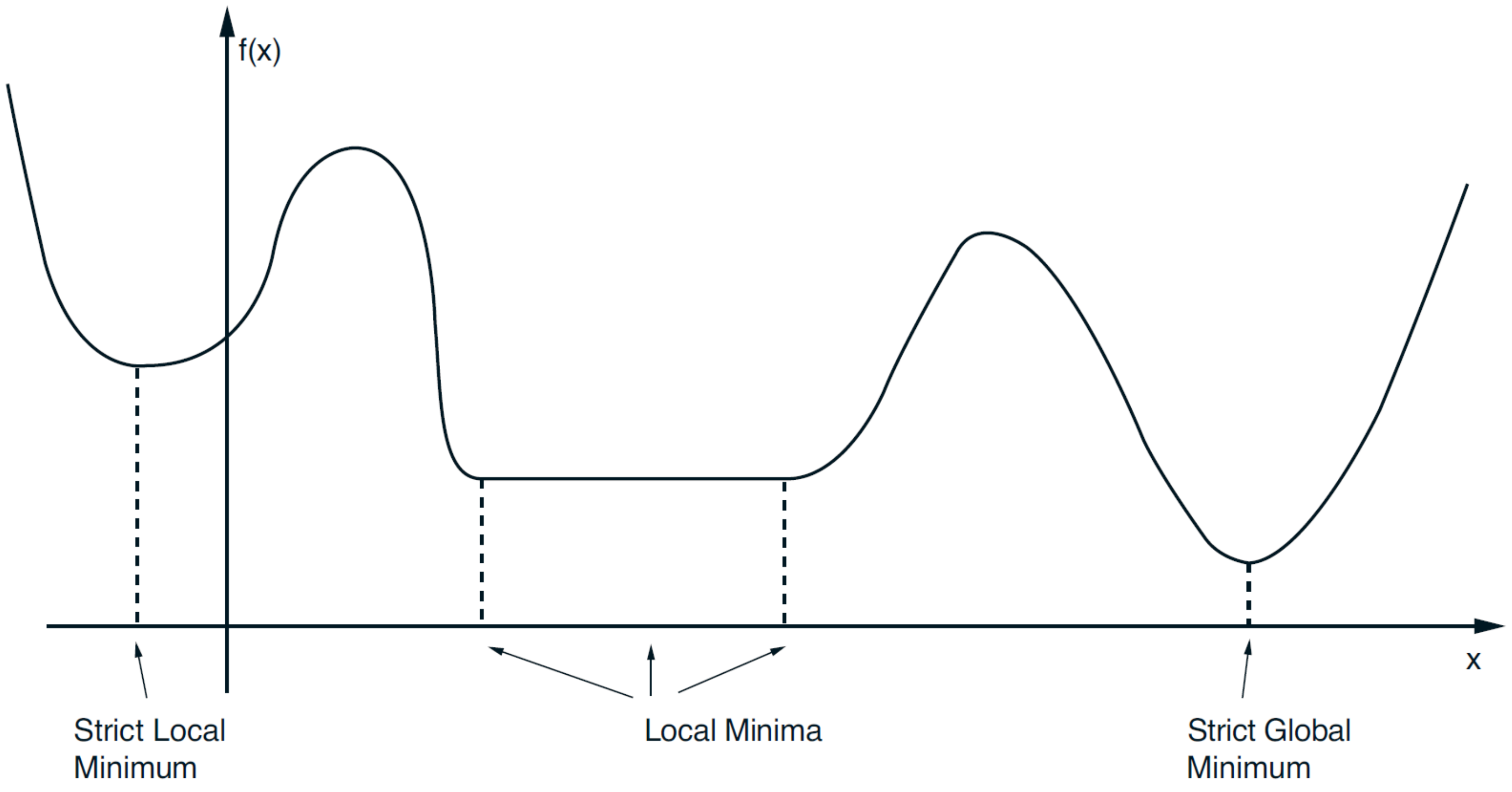
Si X=$\Re^n$ se habla de problemas sin restricciones.

Si f es lineal y X es un poliedro, entonces es un problema de optimización lineal.

- La función objetivo puede tener un solo mínimo, en cuyo caso se denomina unimodal, o varios mínimos locales o globales, en cuyo caso se denomina multimodal

Optimización sin restricciones

Caracterización de los extremos

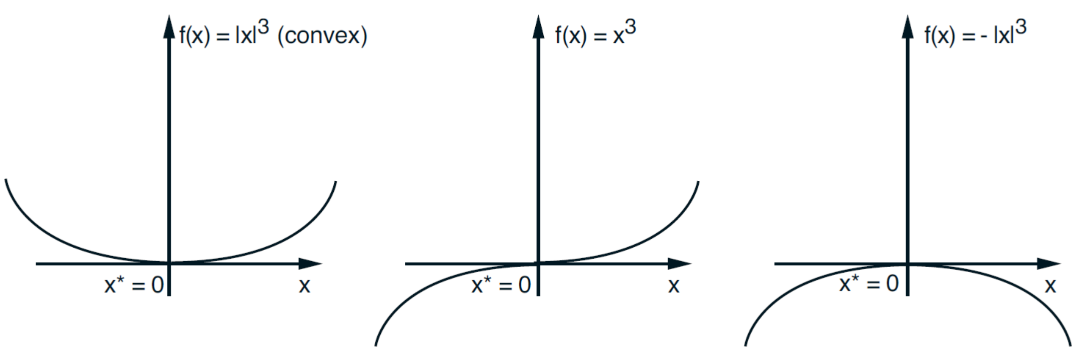


Condición necesaria de optimalidad

- Sea x* un mínimo local de $f:\Re^n \to \Re$ asumimos que f es continuamente diferenciable en un conjunto abierto S que contiene a x*. Entonces,

$$\nabla f(x^*) = 0$$

- Si además f es dos veces continuamente diferenciable en S, entonces, $\nabla^2 f(x^*):$ semidefinida positiva



Condición suficiente de optimalidad

- Sea $f:\Re^n \to \Re$ dos veces continuamente diferenciable. Supóngase que x* satisface:

$$\nabla f(x^*) = 0 \qquad \nabla^2 f(x^*): \text{ definida positiva}$$

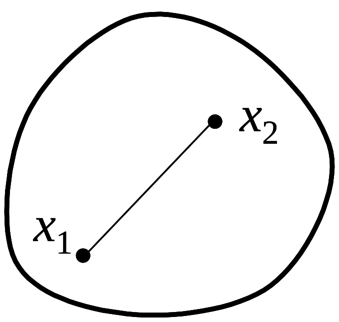
entonces x* es un mínimo local estricto de f.

- Prueba: Sea λ >0 el menor valor propio de $\nabla^2 f(x^*)$

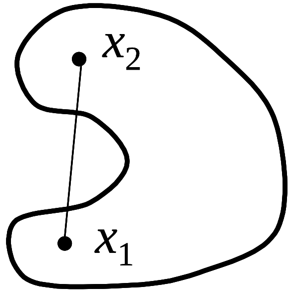
$$\begin{aligned} f(x^* + d) - f(x^*) &= \nabla f(x^*)'d + \frac{1}{2} d'\nabla^2 f(x^*)d \\ &\quad + o(\|d\|^2) \\ &\geq \frac{\lambda}{2}\|d\|^2 + o(\|d\|^2) \\ &= \left(\frac{\lambda}{2} + \frac{o(\|d\|^2)}{\|d\|^2}\right)\|d\|^2. \end{aligned}$$

Repaso: Funciones Convexas

- Estamos particularmente interesados en la optimización de este tipo de funciones sobre los llamados *conjuntos convexos*
- ***Definición***: Un conjunto X en R^n es convexo si y sólo si para dos puntos cualquiera x_1 y x_2 en X y cualquier valor escalar $0 \leq \lambda \leq 1$, el punto $x = \lambda x_1 + (1 - \lambda) x_2$ también está dentro de X



Convexo



No Convexo

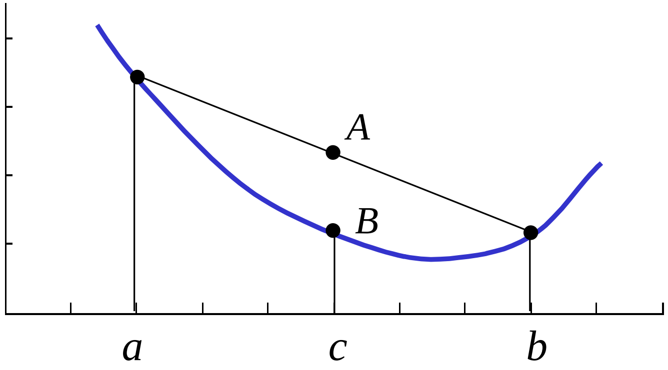
Una esfera, un triángulo, el espacio R^n, una línea recta y un punto son conjuntos convexos. Un hiperplano también es un conjunto convexo

Repaso: Funciones Convexas

- ***Definición***: Una función $f(x)$ es una función convexa definida sobre un conjunto convexo X en R^n si para dos puntos cualquiera x_1 y x_2 en X

$$f\left(\lambda \vec{x}_1 + (1-\lambda)\vec{x}_2\right) \leq \lambda f\left(\vec{x}_1\right) + (1-\lambda) f\left(\vec{x}_2\right)$$

donde $0 \leq \lambda \leq 1$



$$\overline{AC} = \lambda f(a) + (1-\lambda) f(b)$$
$$\overline{BC} = f\left(\lambda a + (1-\lambda) b\right)$$

Repaso: Funciones Convexas

- Las funciones convexas tienen una caracterización geométrica simple e informativa
- ***Teorema*: Cualquier función lineal $f(x) = c^{T}x$ es tanto cóncava como convexa
- ***Teorema***: Si $f(x)$ es convexa → $-f(x)$ es cóncava (y viceversa)
- ***Teorema***: La suma de 2 o más funciones convexas es convexa
- **Teorema:** Cualquier forma cuadrática semidefinida positiva $q(x) = x^{T}Dx$ donde D es simétrica, es una función convexa en todo R^{n}, y si D es definida positiva es estrictamente convexa
- ***Teorema***: Cualquier forma cuadrática semidefinida negativa $q(x) = x^{T}Dx$ donde D es simétrica, es una función cóncava en todo R^{n}, y si D es definida positiva es estrictamente cóncava

Criterios de convexidad

- ***Teorema**:* Supongamos que $f(x)$ tiene primeras derivadas parciales continuas. Luego $f(x)$ es cóncava sobre alguna región R en R^n si y sólo si

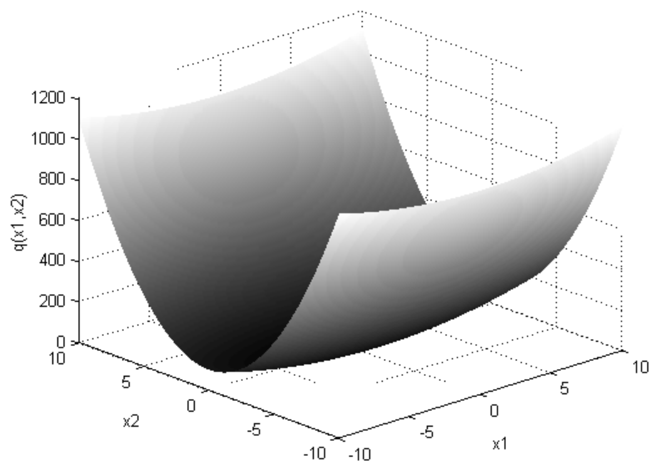
$$f\left(\vec{x}\right) \leq f\left(\vec{x}^{\,*}\right) + \nabla f^{\,t}\left(\vec{x}^{\,*}\right) \cdot \left(\vec{x} - \vec{x}^{\,*}\right)$$

similarmente, $f(x)$ es convexa sobre alguna región R en R^n si y sólo si

$$f\left(\vec{x}\right) \geq f\left(\vec{x}^{\,*}\right) + \nabla f\left(\vec{x}^{\,*}\right)^{t}\left(\vec{x} - \vec{x}^{\,*}\right)$$

Criterios de convexidad

- ***Teorema*:** Sea $f(x)$ una función $\in C^2$ (segundas derivadas parciales existen y son continuas). Entonces $f(x)$ es convexa sobre una región R en R^n si y sólo si su Hessiano es definido o semidefinido positivo para todo x de la región R



Criterios de convexidad

- ***Teorema de Schwartz***: Si $f(x,y)$ es tal que $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ y $\frac{\partial f^{2}}{\partial x \partial y}$ son continuas en un entorno de un punto (x_0,y_0), entonces $\frac{\partial f^{2}}{\partial y \partial x}(x_0 y_0)$ existe y se cumple que

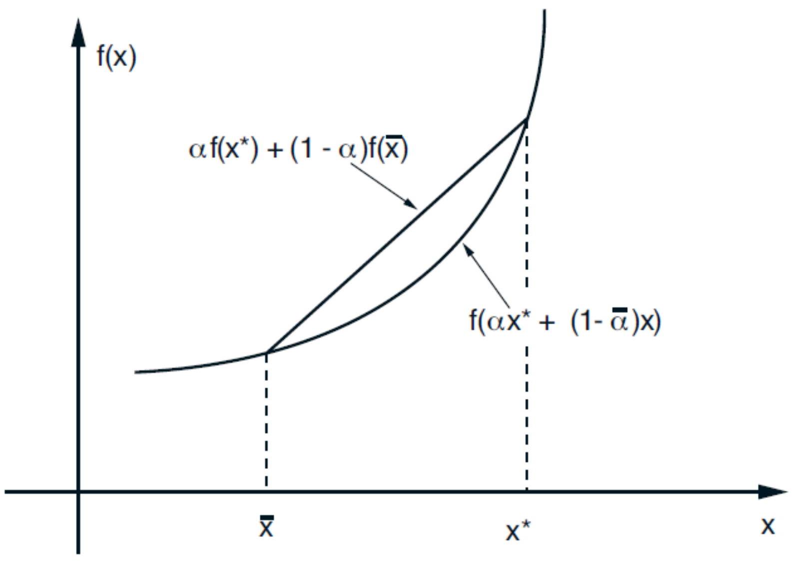
$$\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$$

Criterio de optimalidad: Caso convexo

- Sea f es una función convexa
  - A) Un mínimo local de f, es también un mínimo global. Además, si f es estrictamente convexa entonces existe a lo sumo un mínimo global de f.
  - B)Si f es convexa, entonces

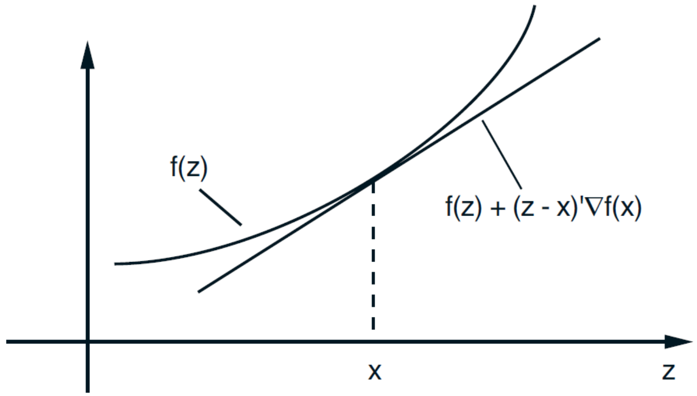
$$\nabla f(x^*) = 0$$

Es una condición necesaria y suficiente para que el vector x* sea un mínimo global de f sobre X.



Optimalidad en el caso convexo

- Se cumple que $f(z) \geq f(x) + \nabla f(x)'(z - x), \qquad \forall\ z \in \Re^n$



Por lo tanto $\nabla f(x^*) = 0$ implica que x* es un mínimo global

Caso cuadrático:

$$\min_{x \in \Re^n} f(x) = \frac{1}{2} x' Q x - b' x$$ Con Q simétrica nxn y b en $\mathcal{R}^n$

Ejemplo: $f(x,y) = \frac{1}{2}\left(\alpha x^2 + \beta y^2\right) - x$

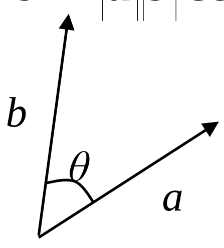
Métodos de descenso para funciones de varias variables

- La forma general de los métodos básicos de descenso se puede expresar como

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k d_k$$

Determinación de Direcciones de Descenso

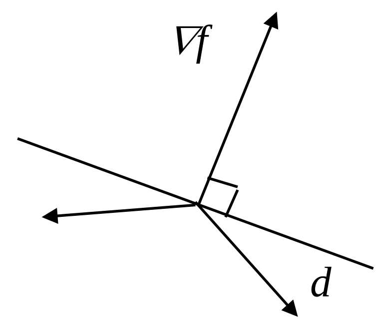
$$\vec{a} \cdot \vec{b} = |a||b| \cos\theta$$



Si $\theta > 90º$, entonces $cos(\theta) < 0$

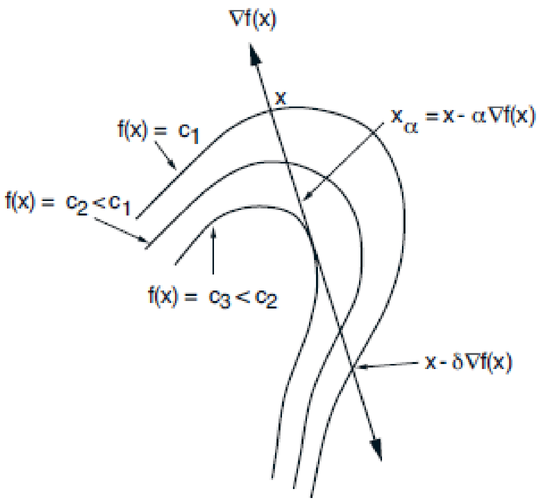
Si sabemos que el gradiente es la máxima dirección de crecimiento en un punto

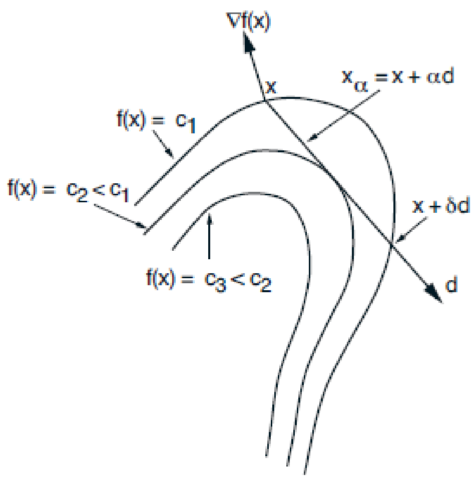
Cualquier vector que tenga más de 90º con el gradiente define una dirección de descenso



$$\vec{d}_K \nabla f < 0$$

Método de descenso





- Si $\nabla f(x) \neq 0$ existe un intervalo de pasos (0,δ) tales que

$$f(x - \alpha \nabla f(x)) < f(x)$$
$$\forall \alpha \in (0, \delta)$$

- Si d tiene un ángulo con $\nabla f(x)$ que es mayor que 90º,

$$\nabla f(x)' d < 0$$

entonces hay un intervalo $(0, \delta)$ de pasos para los cuales

$$f(x + \alpha d) < f(x)$$
$$\forall \alpha \in (0, \delta)$$

Métodos de descenso

Los métodos básicos de descenso entonces son de la forma:

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k d_k$$

donde, si $\nabla f(x) \neq 0$ las direcciones d_k satisfacen:

$$\nabla f(x)' d_k < 0$$

y α_k es un paso positivo.

Formulación de muchos de los métodos:

$$\vec{x}_{k+1} = \vec{x}_k - \alpha_k D_k \nabla f(x_k)$$

donde D_k es una matriz simétrica definida positiva.

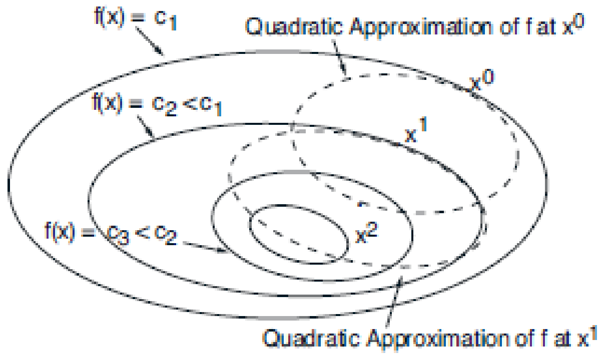
$\nabla f(x)' d_k = -\nabla f(x)' D_k \nabla f(x_k) < 0$ por ser definida positiva

Dos métodos de descenso

- El método de descenso "más rápido" (steepest) (es común llamarlo método del gradiente)

$$\vec{x}_{k+1} = \vec{x}_k - \alpha_k \nabla f(x_k)$$

Es un método de convergencia lenta



- Método de Newton

converge más rápido

$$\vec{x}_{k+1} = \vec{x}_k - \alpha_k (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

Existen muchas modificaciones y variaciones de estos métodos

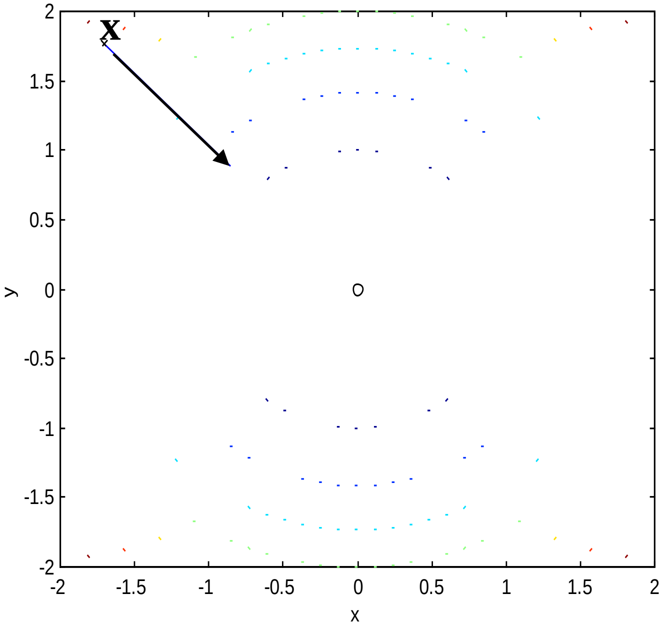
Método del descenso más Rápido

- Ejemplo 1: Se desea minimizar la función $f(x,y) = -20 + x^2 + y^2$

El mínimo está ubicado en el punto (0,0)

Supongamos que se asume como punto inicial, el punto (-1.7, 1.7635)

El gradiente en un punto cualquiera es,
$\nabla f = \{2x, 2y\}$

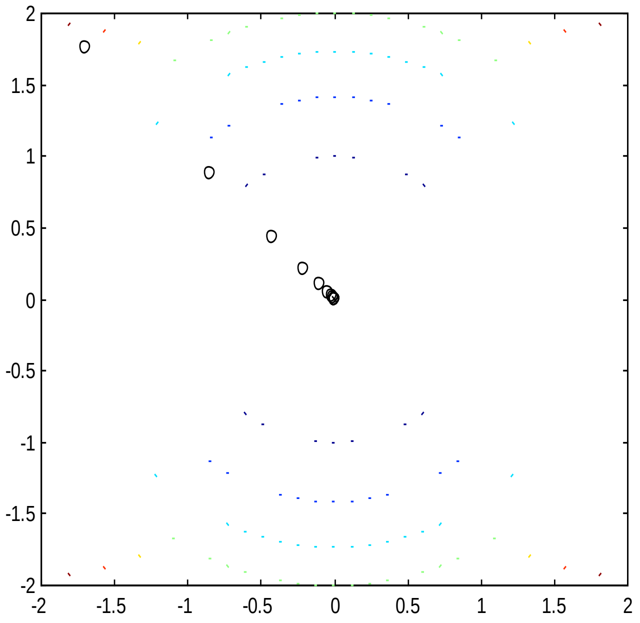


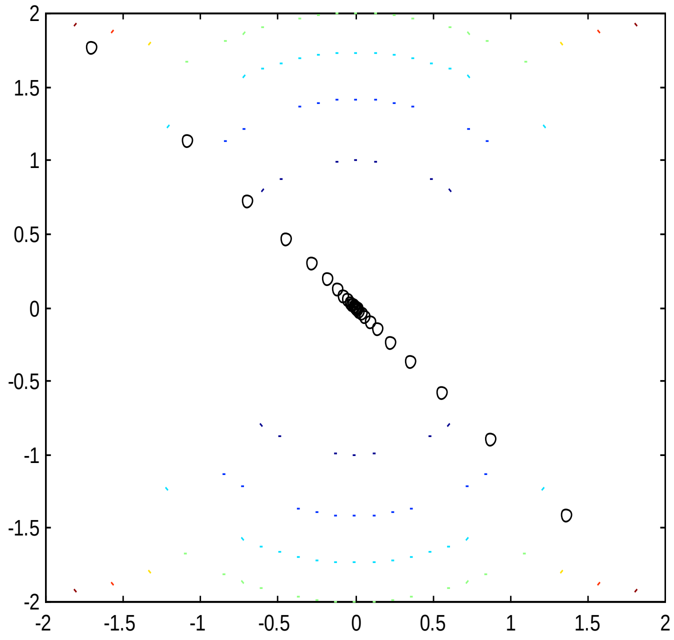
Método del descenso más Rápido

- Ejemplo 1:

Evolución del método para un $\alpha = 0.25$

Evolución del método para un $\alpha = 0.9$





Método del Descenso más Rápido

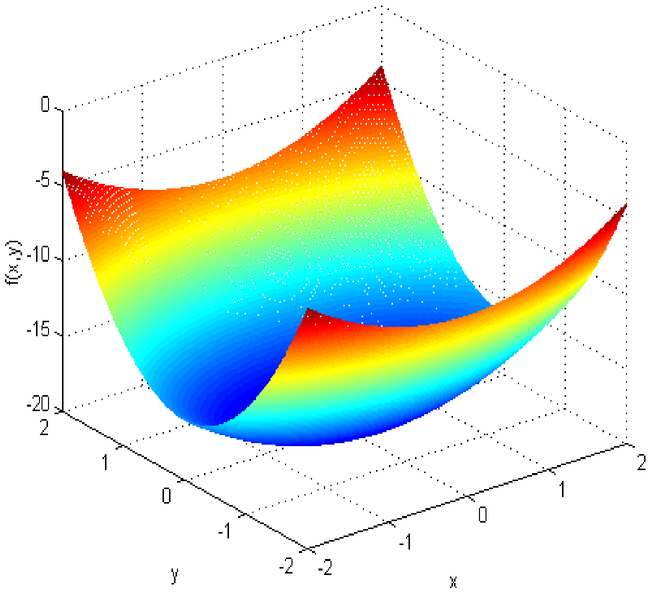
- Ejemplo 2: Se desea minimizar la función $f(x,y) = -20 + 3x^2 + y^2$

Esta función es unimodal

El mínimo está ubicado en el punto (0,0)

Supongamos que se asume como punto inicial, el punto (-1.7, 1.7)

El gradiente en un punto cualquiera es,
$\nabla f = \{6x, 2y\}$

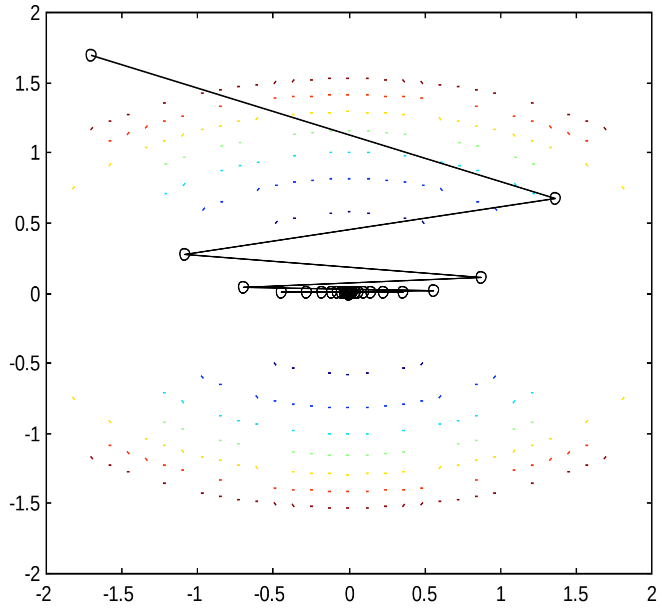


Método del Descenso más Rápido

- Ejemplo 2: Se desea minimizar la función

$$f(x,y) = -20 + 3x^2 + y^2$$

Las curvas de nivel de esta función son de forma elíptica, y el cambio de la dirección de búsqueda de una iteración a otra, se observa en la trayectoria en forma de zigzag



Método de Newton

- Consideramos una aproximación cuadrática de *f(x)* a partir del desarrollo de Taylor:

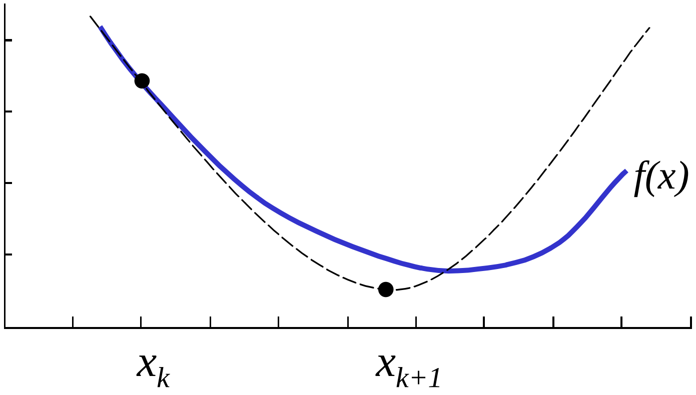
$$f^{k}(x) = f(x_k) + \nabla f(x_k)' \cdot (x - x_k) + \frac{1}{2}(x - x_k)\nabla^2 f(x_k)(x - x_k)$$

- Se puede estimar x_{k+1} determinando el punto donde el gradiente de $f^k(x)$ se hace cero

$$\nabla f^{k}(x) = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0$$

$$x_{k+1} = x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$$

$$x_{k+1} = x_k - \alpha_k \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$$



Método de Newton

- Ejemplo 3: Se desea minimizar la función

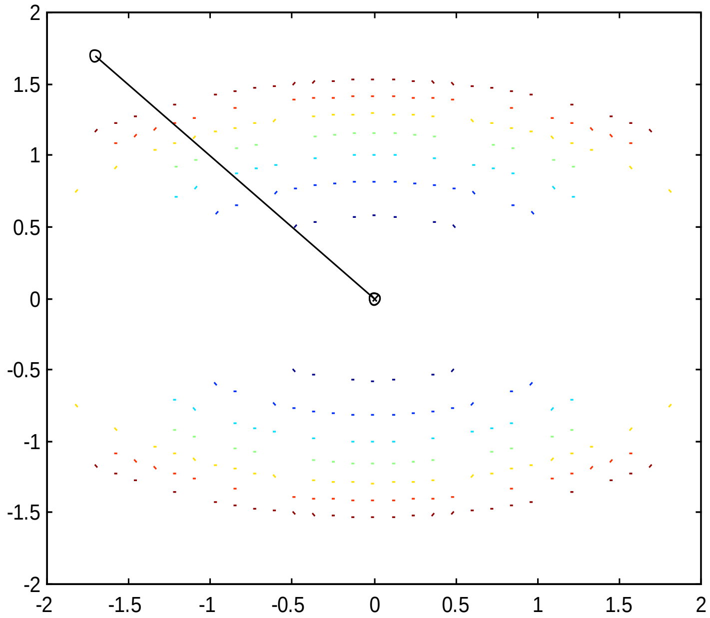
$$f(x,y) = -20 + 3x^2 + y^2$$

utilizando el método de Newton

El gradiente en un punto cualquiera es,
$\nabla f = \{6x, 2y\}$

mientras que el Hessiano es la matriz

$$\boldsymbol{H}_f = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow \boldsymbol{H}_f^{-1} = \begin{bmatrix} 1/6 & 0 \\ 0 & 1/2 \end{bmatrix}$$



La aproximación de esta función utilizando la serie de Taylor es exacta, debido a que es una función cuadrática

Criterios de selección del paso

- Criterio de minimización: $$f(x^k + \alpha^k d^k) = \min_{\alpha \geq 0} f(x^k + \alpha d^k).$$

Hay que resolver otro problema de minimización pero esta vez en una sola variable.

- Minimización limitada. Igual al anterior pero con $\alpha \in [0, s]$

- Paso constante $\alpha^k = s$
- Paso decreciente y verificando la condición:

$$\alpha^k \to 0 \qquad \sum_{k=0}^{\infty} \alpha^k = \infty$$

Existen otros criterios, como por ejemplo criterios de tipo “backtracking”.

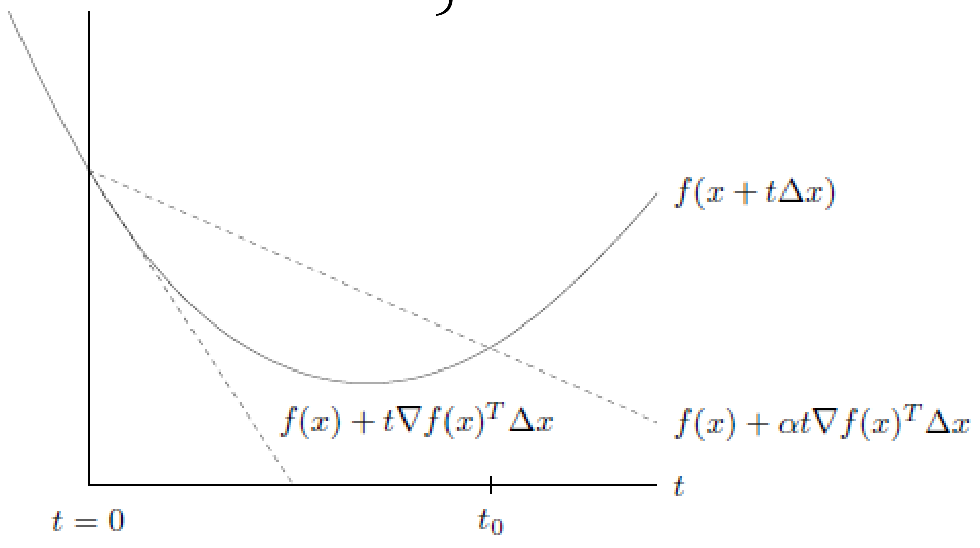
Criterios tipo "backtracking"

Sea d_k una dirección de descenso,

y sean dados $\beta \in [0,1]$ y $\alpha \in [0, 0.5]$

$t := 1$

$$while \left\{ \begin{array}{l} f(x_k + td_k) \; > \; f(x_k) + \alpha t \nabla' f(x_k) d_k \\ t =: \beta t \end{array} \right\}$$



Convergencia

Por un lado, debe existir una disminución significativa de f:

$$\frac{\nabla'(f(x_k))d_k}{\|\nabla(f(x_k))\|\|d_k\|}$$

si

$$d_k = -D_k \nabla(f(x_k))$$

se puede probar que si los valores propios de las matrices D_k están acotados superior e inferiormente por reales positivos

$$m\|z\|^2 \leq z' D_k z \leq M\|z\|^2$$

$$\nabla'(f(x_k))d_k = \nabla'(f(x_k))D_k\nabla(f(x_k)) \geq m\|\nabla(f(x_k))\|^2$$

$$\|d_k\|^2 = \nabla'(f(x_k))D_k{}^2\nabla(f(x_k)) \leq M^2\|\nabla(f(x_k))\|^2$$

Convergencia

- Además de la dirección de descenso importa elegir el paso adecuadamente.
- Por lo tanto las pruebas de convergencia dependen de que la dirección de descenso sea “buena” y que se elija un “buen” paso de iteración.

Teoremas de convergencia

1) Sea la secuencia $x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$

Si el paso α_k se elije mediante un método de minimización o por backtracking y la matriz D_k verifica la condición vista para las cotas de sus valores propios, entonces todo punto límite de $\{x_k\}$ es estacionario.

Existen resultados más generales que la condición en los valores propios.

Teoremas de convergencia Paso constante

2) Sea la secuencia $x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$ con $d_k = -D_k \nabla f(x_k)$

Si la matriz D_k verifica la condición vista para las cotas de sus valores propios, se asume además que:

$$\text{Existe } L > 0 \;\; tal\;\; que \;\; \left\| \nabla f\left(x\right) - \nabla f\left(y\right) \right\| \leq L \left\| x - y \right\| \quad \forall x, y \;\; (Lipschitz)$$

$$\forall k \quad d_k \neq 0 \text{ y } \varepsilon < \alpha_k < (2 - \varepsilon)\beta \;\; \text{con } \beta = \frac{\left| \nabla f\left(x_k\right) d_k \right|}{L \left\| d_k \right\|}$$

$$\varepsilon > 0 \text{ fijo.}$$

entonces todo punto límite de $\{x_k\}$ es estacionario.

Velocidad de convergencia

- Se analiza a través de una función de error. Las más comúnmente utilizadas son:

$$e(x) = \|x - x^*\|, \qquad e(x) = f(x) - f(x^*)$$

- Convergencia geométrica o lineal: $e(x^k) \leq q\beta^k \quad q > 0 \quad \beta \in [0,1)$,

$$\limsup_{k\to\infty} \frac{e(x^{k+1})}{e(x^k)} < \beta$$

- Convergencia superlineal $e(x^k) \leq q \cdot \beta^{p^k} \quad q > 0, p > 1$

$$\limsup_{k\to\infty} \frac{e(x^{k+1})}{e(x^k)} = 0$$

- Convergencia sublineal

Velocidad de convergencia del método de gradiente

- Se puede probar que utilizando el método de minimización del paso
$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \leq 1 - \frac{m}{M}$$
asumiendo que existen m y M el menor y mayor valor propio de
$\nabla^2(f(x_k)) \quad \forall k$

- Es decir que la convergencia es lineal.

Velocidad de convergencia del método de Newton

- Idea: $x^{k+1} = x^k - \left(\nabla g(x^k)'\right)^{-1} g(x^k)$, observar que si $g(x) = \nabla f(x)$ se tiene la forma pura del algoritmo de Newton.

Si $x^k \;\rightarrow\; x^*$ con $g(x^*) = 0$ y el gradiente de g(x) es invertible,

$$0 = g(x^*) = g(x^k) + \nabla g(x^k)'(x^* - x^k) + o\left(\|x^k - x^*\|\right)$$

$$x^k - x^* - \left(\nabla g(x^k)'\right)^{-1} g(x^k) = o\left(\|x^k - x^*\|\right)$$

$$x^{k+1} - x^* = o\left(\|x^k - x^*\|\right)$$

Lo cual implica convergencia superlineal