Condiciones de Optimalidad

Alan Reyes-Figueroa

Métodos Numéricos II (Aula 15) 23.agosto.2021

Definición

Suponga que $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ *es una función con valaores reales, definida sobre* Ω. *Un punto* $\mathbf{x}^* \in \Omega$ *es un* ***mínimo local*** *o* ***minimizador local*** *de* f *si existe* $\varepsilon > 0$ *tal que*

$$f(\mathbf{x}) \geq f(\mathbf{x}^*), \qquad \textit{para todo } \mathbf{x} \in \Omega - \{\mathbf{x}^*\} \textit{ con } ||\mathbf{x} - \mathbf{x}^*|| < \varepsilon.$$

Definición

Suponga que $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ *es una función con valaores reales, definida sobre* Ω. *Un punto* $\mathbf{x}^* \in \Omega$ *es un* ***mínimo global*** *o* ***minimizador global*** *de* f *sobre* Ω *si*

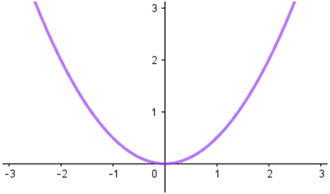
$$f(\mathbf{x}) \geq f(\mathbf{x}^*), \qquad \textit{para todo } \mathbf{x} \in \Omega, \ \mathbf{x} \neq \mathbf{x}^*.$$

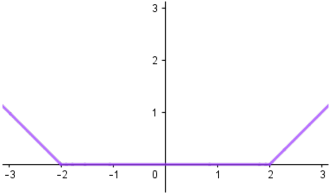
Obs! Reemplazando $\geq$ con $>$ en las definiciones anteriores obtenemos el concepto de un **mínimo local estricto** y de un **mínimo global estricto**, respectivamente.

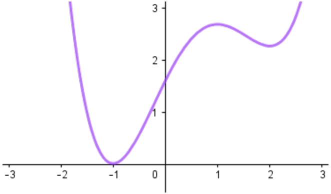
Ejemplo: La función $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ tiene un mínimo global estricto en $x = 0$. Claramente, $x = 0$ también es un mínimo local de f.

Ejemplo: La función $f : \mathbb{R} \to \mathbb{R}$, dada por $f(x) = \max\{0, |x - 2|\}$ tiene a todos los puntos de intervalo $[-2, 2]$ como mínimos globales. Estos no son estrictos.

Ejemplo: La función $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 3x^4 - 8x^3 - 6x^2 + 12x$ tiene mínimos locales estricto en $x = -1$ y $x = 2$. De éstos, sólo $x = -1$ es un mínimo global.





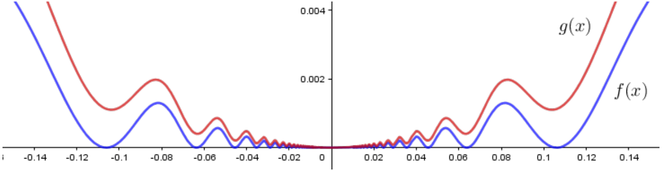


Definición

Un punto $\mathbf{x}^* \in \Omega$ *es un* ***mínimo local aislado*** *de* $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, *si* $\mathbf{x}^*$ *es mínimo local de* f *y existe una vecindad* $U \subset \mathbb{R}^n$ *de* $\mathbf{x}^*$ *tal que* $\mathbf{x}^*$ *es el único mínimo local de* f *en* U.
Un punto $\mathbf{x}^* \in \Omega$ *es un* ***mínimo local no aislado*** *de* f, *si para toda vecindad* U *de* $\mathbf{x}^*$, *existe* $\mathbf{x} \in U$, $\mathbf{x} \neq \mathbf{x}^*$, *tal que* $\mathbf{x}$ *también es mínimo local de* f.

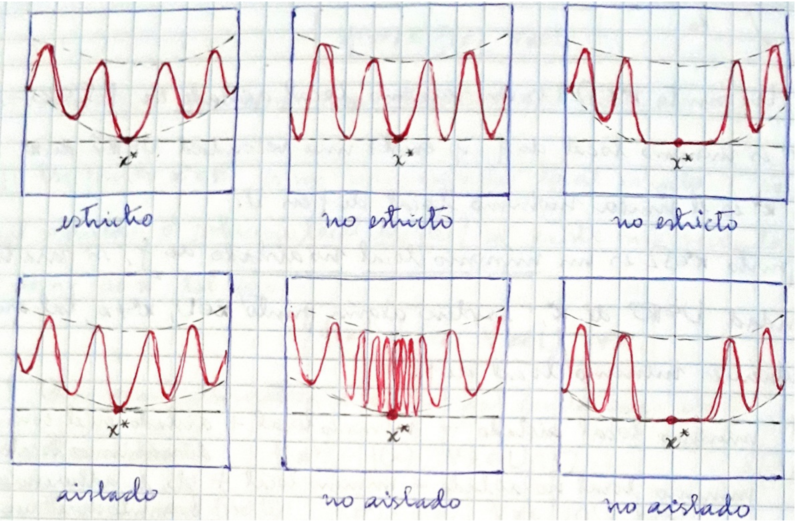
Ejemplo: La función $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2 \cos \frac{1}{x} + x^2$, $f(0) = 0$, posee un mínimo local no estricto y no aislado en $x = 0$.

Ejemplo: La función $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^2 \cos \frac{1}{x} + 2x^2$, $g(0) = 0$, posee un mínimo local estricto y no aislado en $x = 0$.



UNIVERSIDAD DEL VALLE DE GUATEMALA

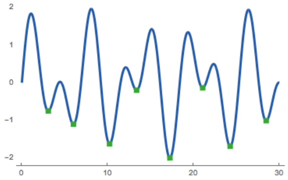
Optimización

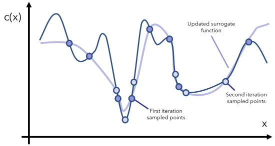


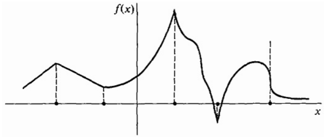
Optimización

Dificultades de la optimización global:

- funciones con muchos mínimos;
- los algoritmos tienden a quedarse atrapados en mínimos locales,
- en el caso de métodos de búsqueda, puede que el óptimo global esté en una región no explorada;
- cuando el mínimo se encuentra dentro de una región donde la función es muy plana (curvatura cercana a 0), los métodos de optimización suelen ser muy lentos;
- no-diferenciabilidad en un punto mínimo.







Optimización

Fórmula de Taylor:
Si $f : \mathbb{R}^n \to \mathbb{R}$ es de clase C^{m+1} sobre $\mathbb{R}^n$, y $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$, entonces

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{k=0}^{m} \frac{1}{k!} D^{(k)} f(\mathbf{x}_0) \cdot \mathbf{h}^{(k)} + \frac{1}{(m+1)!} D^{(m+1)} f(\mathbf{x}_0 + t\mathbf{h}) \cdot \mathbf{h}^{(m+1)},$$

donde $t \in (0,1)$ y

$$D^{(k)} f(\mathbf{x}_0) \cdot \mathbf{h}^k = \sum_{|I|=k} \frac{\partial^k f}{\partial \mathbf{x}_I}(\mathbf{x}_0)\, \mathbf{h}_I = \sum_{|I|=k} \frac{\partial^k f}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}}\, h_1^{i_1} h_2^{i_2} \cdots h_n^{i_n},$$

$I = (i_1, \ldots, i_n)$, $\mathbf{x}_I = (x_{i_1}, \ldots, x_{i_n})$, $\mathbf{h}_I = (h_{i_1}, \ldots, h_{i_n})$.

Casos particulares:
Si $f : \mathbb{R}^n \to \mathbb{R}$ es de clase C^2, podemos escribir

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{x}_0) + Df(\mathbf{x}_0 + t\mathbf{h}), \quad t \in (0,1). \\ f(\mathbf{x}) &= f(\mathbf{x}_0) + Df(\mathbf{x}_0) \cdot \mathbf{h} + \tfrac{1}{2}\, \mathbf{h}^T D^2 f(\mathbf{x}_0 + t\mathbf{h})\, \mathbf{h}, \quad t \in (0,1). \end{aligned}$$

Teorema (Condiciones de Optimalidad de Primer Orden)

Si $\mathbf{x}^*$ es un mínimo local (o un máximo local) de la función $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, y f es de clase C^1 en una vecindad abierta de $\mathbf{x}^*$, entonces $\nabla f(\mathbf{x}^*) = \mathbf{0}$.

Prueba: Hacemos la prueba para $\mathbf{x}^*$ mínimo local. Suponga que $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$. Por lo tanto, podemos encontrar una dirección $\mathbf{h} = -\alpha \frac{\nabla f(\mathbf{x}^*)}{||\nabla f(\mathbf{x}^*)||} = -\alpha \mathbf{u}$, tal que
$D_{\mathbf{u}} f(\mathbf{x}^*) = \nabla f(\mathbf{x}^*)^T \mathbf{u} < 0$.

Por el Teorema de Taylor, si $\mathbf{x} = \mathbf{x}^* + \mathbf{h}$, tenemos $f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T \mathbf{h} + o(||\mathbf{h}||)$. Cuando $\alpha \to 0$, entonces $\mathbf{h} \to 0$, resulta que $\nabla f(\mathbf{x}^*)^T \mathbf{h} + o(||\mathbf{h}||) < 0$, ya que $o(||\mathbf{h}||)$ se acerca a cero, mucho más rápido que $\nabla f(\mathbf{x}^*)^T \mathbf{h}$. De hecho,

$$\lim_{\alpha \to 0} \frac{|\nabla f(\mathbf{x}^*)^T \mathbf{h}|}{||\mathbf{h}||} = \frac{\nabla f(\mathbf{x}^*)^T \mathbf{u}}{||\mathbf{u}||} = \nabla f(\mathbf{x}^*)^T \mathbf{u} < 0.$$

De ahí que, $f(\mathbf{x}) < f(\mathbf{x}^*)$, y esto contradice el la hipótesis de que $\mathbf{x}^*$ es un mínimo local. Portanto, $\nabla f(\mathbf{x}^*) = \mathbf{0}$. $\square$

Definición

*Un punto $\mathbf{x} \in \Omega = \operatorname{dom} f$ que satisface que $\nabla f(\mathbf{x}) = \mathbf{0}$ se llama un **punto crítico** o **punto estacionario** de f.*

De acuerto a la condición de Optimalizad de Primer Orden, todo mínimo local debe ser un punto estacionario de f.

Teorema (Condiciones de Optimalidad de Segundo Orden)

Si $\mathbf{x}^* \in \mathbb{R}^n$ es un mínimo (máximo) localde $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, y f es de clase C^2 es una vecindad abierta de $\mathbf{x}^*$, entonces $\nabla f(\mathbf{x}^*) = \mathbf{0}$, y la hessiana

$$D^2 f(\mathbf{x}^*) \succeq 0$$

es positiva semidefinida (negativa semidefinida).

Prueba: Al igual que antes, hacemos la prueba para el caso de $\mathbf{x}^*$ mínimo local. El caso del máximo se prueba de forma similar.

Condiciones de Optimalidad

Suponga que $D^2f(\mathbf{x}^*)$ no es positiva semidefinida. Entonces, existe $\mathbf{h} \in \mathbb{R}^n$ tal que $\mathbf{h}^T\, D^2f(\mathbf{x}^*)\,\mathbf{h} < 0$.
Por continuidad de D^2f, y la preservación de signo, existe una bola $\mathbb{D}_r(\mathbf{x}^*)$ y un intervalo $(0,\varepsilon)$, tal que $\mathbf{h}^T\, D^2f(\mathbf{x}^* + \widehat{\varepsilon}\mathbf{h})\,\mathbf{h} < 0$, para todo $\widehat{\varepsilon} \in (0,\varepsilon)$.

Aplicando el Teorema de Taylor alrededor de $\mathbf{x}^*$, existe $t \in (0,1)$, tal que

$$f(\mathbf{x} + \widehat{\varepsilon}\mathbf{h}) = f(\mathbf{x}^*) + \widehat{\varepsilon}\nabla f(\mathbf{x}^*)^T\mathbf{h} + \tfrac{1}{2}\widehat{\varepsilon}^2\mathbf{h}^T\, D^2f(\mathbf{x}^* + t\widehat{\varepsilon}\mathbf{h})\,\mathbf{h}.$$

Usando el hecho que $\nabla f(\mathbf{x}^*)^T\mathbf{h} = \mathbf{0}$ (ya que $\mathbf{x}^*$ es mínimo local), y el hecho que $\mathbf{h}^T\, D^2f(\mathbf{x}^* + t\widehat{\varepsilon}\mathbf{h})\,\mathbf{h} < 0$, obtenemos

$$f(\mathbf{x}^* + \widehat{\varepsilon}\mathbf{h}) < f(\mathbf{x}^*),$$

lo cual contradice la hipótesis de que $\mathbf{x}^*$ es un mínimo local de f. Portanto, $D^2f(\mathbf{x}^*)$ es positiva semidefinida. $\square$

Teorema (Condiciones Suficientes de Optimalidad)

Suponga que $D^2 f$ existe y es continua en una vecindad de $\mathbf{x}^* \in \mathbb{R}^n$, que $\nabla f(\mathbf{x}^*) = \mathbf{0}$, y que la hessiana $D^2 f(\mathbf{x}^*)$ es positiva definida (negativa definida).
Entonces $\mathbf{x}^*$ es un mínimo (máximo) local estricto de f.

Prueba: Existe una bola $\mathbb{D}_r(\mathbf{x}^*) = \{\mathbf{x} : ||\mathbf{x} - \mathbf{x}^*|| < r\}$ para la cual
$q(t) = \mathbf{h}^T D^2 f(\mathbf{x}^* + t\mathbf{h})\,\mathbf{h} > 0$, para todo $\mathbf{x}^* + t\mathbf{h} \in \mathbb{D}_r$, con $t \in (0,1)$ y $\mathbf{h} = \mathbf{x} - \mathbf{x}^*$.
(Esto es consecuencia de la preservación de signo, ya que la función q es continua y $q(0) = \mathbf{h}^T D^2 f(\mathbf{x}^*)\,\mathbf{h} > 0$).

Usando el Teorema de Taylor, con $||\mathbf{h}|| < r$, y como $\nabla f(\mathbf{x}^*) = \mathbf{0}$, se tiene que existe $t \in (0,1)$ tal que

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \tfrac{1}{2}\mathbf{h}^T D^2 f(\mathbf{x}^* + t\mathbf{h})\,\mathbf{h},$$

Como $\mathbf{x}^* + t\mathbf{h} \in \mathbb{D}_r(\mathbf{x}^*)$, luego $\mathbf{h}^T D^2 f(\mathbf{x}^* + t\mathbf{h})\,\mathbf{h} > 0 \;\Rightarrow\; f(\mathbf{x}) > f(\mathbf{x}^*)$, para todo $\mathbf{x} \in \mathbb{D}_r(\mathbf{x}^*)$. Esto muestra que $\mathbf{x}^*$ es un mínimo local estricto de f. $\square$

Condiciones de Optimalidad

Podemos encontrar y clasificar puntos estacionarios de la siguiente manera:

1. Encontrar los puntos críticos $\mathbf{x}$ en los que $f(\mathbf{x}) = 0$.
2. Obtener la Hessiana $Hf(\mathbf{x}) = D^2 f(\mathbf{x})$.
3. Determinar el carácter de $Hf(\mathbf{x})$ para cada punto crítico $\mathbf{x}$.
   - Si $D^2 f(\mathbf{x})$ es positiva (negativa) definida, entonces $\mathbf{x}$ es un mínimo (máximo) local.
   - Si $D^2 f(\mathbf{x})$ es indefinida, $\mathbf{x}$ es un punto silla.
   - Si $D^2 f(\mathbf{x})$ es positiva (negativa) semidefinida, $\mathbf{x}$ puede ser un mínimo (máximo) local. En este caso, es necesario seguir trabajando para clasificar el punto estacionario.

     Un posible enfoque sería para deducir las terceras derivadas parciales de $f(\mathbf{x})$ y luego calcular el término correspondiente en la serie de Taylor. Si este término es cero, entonces el siguiente término necesita ser calculado y así por delante.

Condiciones de Optimalidad

- En el caso especial donde $D^2 f(\mathbf{x}) = 0$, $\mathbf{x}$ puede ser un minimizador o maximizador ya que las condiciones necesarias se satisfacen tanto en casos.
- Si $D^2 f(\mathbf{x})$ es semidefinida, se requiere más información para caracterización completa de un punto estacionario y más el trabajo es necesario en este caso.
- Un posible enfoque podría ser calcular el tercer término de la serie de Taylor de $f(\mathbf{x})$,

$$D^3 f(\mathbf{x}) \cdot \mathbf{h}^{(3)} = \frac{1}{3!} \sum_{|I|=3} \frac{\partial^3 f}{\partial \mathbf{x}_I}(\mathbf{x}_0)\, \mathbf{h}_I = \frac{1}{3!} \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} \frac{\partial^3 f}{\partial x_1^{i_1} \partial x_2^{i_2} \partial x_3^{i_3}}\; h_1^{i_1} h_2^{i_2} h_3^{i_3};$$

  y se debe determinar el signo de este término. Si este término es cero, entonces debe calcularse el siguiente término $D^4 f(\mathbf{x})$.
- En general, si los primeros i términos $D^i f(\mathbf{x})$ de la serie de Taylor son todos nulos, debe calcularse el signo de primer término $D^k f(\mathbf{x}) \cdot \mathbf{h}^{(k)}$ que no se anule.

Condiciones de Optimalidad

Ejemplo: Considere la función $f : \mathbb{R}^2 \to \mathbb{R}$, dada por $f(\mathbf{x}) = \frac{1}{6}[(x_1 - 2)^3 + (x_2 - 3)^3]$. En este caso, el gradiente es

$$\nabla f(\mathbf{x}) = \frac{1}{2}\begin{pmatrix}(x_1 - 2)^2 & (x_2 - 3)^2\end{pmatrix}.$$

Resolviendo $\nabla f(\mathbf{x}) = \mathbf{0}$, se obtiene que el único punto crítico es $\mathbf{x}^* = (2, 3)^T$. La Hessiana de f es

$$D^2 f(\mathbf{x}) = \begin{pmatrix} x_1 - 2 & 0 \\ 0 & x_2 - 3 \end{pmatrix}.$$

Así, en el punto crítico, $D^2(\mathbf{x}^*) = \mathbf{0}$. Las terceras derivadas de f son todas cero, excepto

$$\frac{\partial^3 f}{\partial x_1^3}(\mathbf{x}) = \frac{\partial^3 f}{\partial x_2^3}(\mathbf{x}) = 1.$$

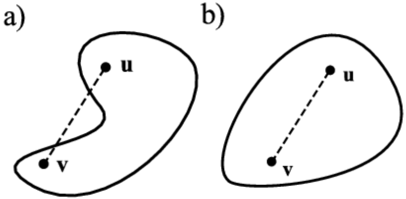
Luego, el término

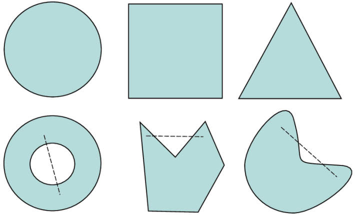
$$D^3 f(\mathbf{x}^*) \cdot \mathbf{h}^{(3)} = \frac{1}{3!} \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} \frac{\partial^3 f}{\partial x_1^{i_1} \partial x_2^{i_2} \partial x_3^{i_3}} \, h_1^{i_1} h_2^{i_2} h_3^{i_3} = \frac{1}{6}(h_1^3 + h_2^3)$$

es positivo si $h_1, h_2 > 0$, pero es negativo si $h_1, h_2 < 0$. Portanto, $D^3 f(\mathbf{x}^*) \cdot \mathbf{h}^{(3)}$ toma ambos signos, y $\mathbf{x}^*$ es un punto silla de f.

Definición

Un subconjunto $\Omega \subseteq \mathbb{R}^n$ *es* ***convexo*** *si para todo* $\mathbf{x}, \mathbf{y} \in \Omega$*, el segmento de recta* $[\mathbf{x}, \mathbf{y}] = \{(1-t)\mathbf{x} + t\mathbf{y} : t \in [0,1]\}$ *está totalmente contenido en* Ω*.*





(a) Conjunto no convexo, (b) Conjunto convexo.

Ejemplos:

- Convexos: esferas, hiperplanos, semiespacios, conos, ...
- No Convexos: conjunto no conexos, uniones de rectas, uniones en general, ...

Definición

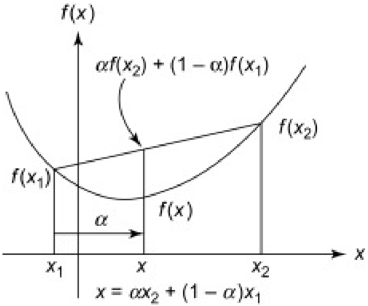
*Una función $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ es **convexa** si $\Omega = \operatorname{dom} f$ es un conjunto convexo, y para todo $\mathbf{x}, \mathbf{y} \in \Omega$, y todo $t \in [0, 1]$ vale*

$$f\big((1-t)\mathbf{x} + t\mathbf{y}\big) \leq (1-t)f(\mathbf{x}) + tf(\mathbf{y}). \tag{1}$$

Geométricamente, la desigualdad (1) significa que el segmento de recta entre $(\mathbf{x}, f(\mathbf{x}))$ y $(\mathbf{y}, f(\mathbf{y}))$ está por encima de la gráfica de f.

La función f es **estrictamente convexa** si en (1) vale la desigualdad estricta, siempre que $\mathbf{x} \neq \mathbf{y}$ y $t \neq 0, 1$. Decimos que f es **cóncava** (**estrictamente cóncava**) si $-f$ es convexa (estrictamente convexa).

A la desigualdad (1) se le llama usualmente **desigualdad de Jensen**.



Propiedad

Sea $\Omega \subseteq \mathbb{R}^n$ conjunto convexo. La función $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ es convexa $\iff$ para todo $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \Omega$, y cualesquiera $t_1, \ldots, t_k \in [0, 1]$, con $\sum_{i=1}^k t_i = 1$, se tiene que

$$f\Big(\sum_{i=1}^{k} t_i \mathbf{x}_i\Big) \leq \sum_{i=1}^{k} t_i f(\mathbf{x}_i). \tag{2}$$

<u>Prueba</u>: ($\Leftarrow$) Para $k = 2$, tome $\mathbf{x}_1 = \mathbf{x}, \mathbf{x}_2 = \mathbf{y} \in \Omega$, y sean $t_1 = 1 - t, t_2 = t$, con $t \in [0, 1]$. La desigualdad (2) se reduce a $f((1 - t)\mathbf{x} + t\mathbf{y}) \leq (1 - t)f(\mathbf{x}) + tf(\mathbf{y})$, lo que implica que f es convexa.

($\Rightarrow$) Mostramos la desigualdad (2) por inducción sobre k.
Para $k = 1$, necesariamente $t_1 = 1$ de modo que $f(\mathbf{x}_1) \leq f(\mathbf{x}_1)$ y (2) se cumple de manera automática. El caso $k = 2$ se cumple a partir de la definición de convexidad (1).

Suponga que (2) se cumple para cualesquiera k puntos $\mathbf{p}_1, \ldots, \mathbf{p}_k \in \Omega$, siempre que se forme una combinación lineal convexa $s_1\mathbf{p}_1 + \ldots + s_k\mathbf{p}_k$, con $0 \leq s_i \leq 1$ y $\sum_{i=1}^k s_i = 1$.

Suponga ahora que $\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1} \in \Omega$, se combinan para formar un punto

$$\mathbf{x} = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \ldots + t_{k+1}\mathbf{x}_{k+1} \in \Omega, \quad \sum_{i=1}^{k+1} t_i = 1, \ 0 \leq t_i \leq 1.$$

Definamos $t = t_{k+1}$, $1 - t = \sum_{j=1}^{k} t_j = t_1 + \ldots + t_k$. Ambos coeficientes satisfacen $0 \leq t, 1 - t \leq 1$. En particular, si $\mathbf{p} = \sum_{j=1}^{k} s_j\mathbf{x}_j \in \Omega$, con $\sum_{j=1}^{k} s_j = 1$, podemos escribir

$$\mathbf{x} = (1-t)\mathbf{p} + t\mathbf{x}_{k+1} = (1-t)\sum_{j=1}^{k} s_j\mathbf{x}_j + t_{k+1} \quad \Longrightarrow \quad t_j = (1-t)s_j, \ j = 1, \ldots, k;$$

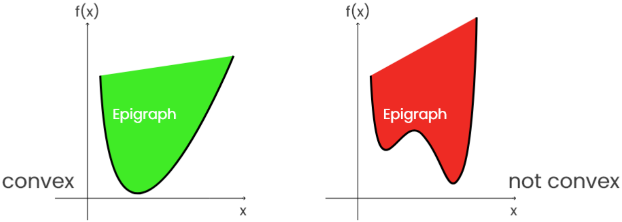
y

$$\begin{aligned}
f\left(\textstyle\sum_{i=1}^{k+1} t_i\,\mathbf{x}_i\right) &= f\big((1-t)\mathbf{p} + t\,\mathbf{x}_{k+1}\big) \ \leq \ (1-t)f(\mathbf{p}) + tf(\mathbf{x}_{k+1}) \\
&\leq (1-t)f\left(\textstyle\sum_{j=1}^{k} s_j\,\mathbf{x}_j\right) + tf(\mathbf{x}_{k+1}) \ \leq \ (1-t)\textstyle\sum_{j=1}^{k} s_j f(\mathbf{x}_j) + tf(\mathbf{x}_{k+1}) \\
&\leq \textstyle\sum_{j=1}^{k} t_j f(\mathbf{x}_j) + tf(\mathbf{x}_{k+1}) \ \leq \ \textstyle\sum_{i=1}^{k+1} t_i f(\mathbf{x}_i). \ \square
\end{aligned}$$

Definición

*Sea $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$. Definimos el **epígrafo** de f, como el conjunto*

$$\mathrm{Epi}(f) = \{(\mathbf{x}, y) \in \mathbb{R}^{n+1} : \; y \geq f(\mathbf{x})\} \subseteq \mathbb{R}^{n+1}.$$



Teorema

f es convexa $\iff$ su epígrafo $\mathrm{Epi}(f)$ *es un conjunto convexo.*