Programación No Lineal: nociones básicas

Jorge López Lázaro

jorloplaz@gmail.com

Enero 2022

1 | Introducción

Introducción

Hemos estudiado los problemas de programación lineal, que modelizan situaciones donde la función objetivo y las restricciones son lineales en las variables de decisión.

Aunque los problemas de programación lineal son muy comunes y cubren un amplio rango de aplicaciones, en la vida real nos enfrentamos a otro tipo de problemas que no son lineales.

Cuando el conjunto de restricciones, la función objetivo, o ambos, son no lineales, se dice que se trata de un problema de **programación no lineal** (NLP).

Introducción

Ejemplos

- Problema de transporte con descuentos por cantidad: El precio por unidad de transporte entre un origen y un destino es decreciente en función de la cantidad a transportar.

- Problema de producción con elasticidad en el precio y/o en el coste: Consideremos que los precios unitarios disminuyen cuando vendemos un número importante de productos, esto nos indica que la renta marginal disminuye cuando aumentan las ventas.
Es el caso de reducir el precio de venta a partir de cierta cantidad vendida: cuando llevo vendida una cantidad importante, le reduzco el precio de venta

Introducción

Problema de transporte con descuentos por cantidad (I)

Una situación muy habitual es que, en el problema del transporte, se disponga de *descuentos por cantidad* para volúmenes grandes, con lo que la función de coste unitaria sería una función no lineal con pendiente no creciente.

Una alternativa es aproximar esta función no lineal por una poligonal.

Así pues, el coste de embarcar unidades viene dado por una función *poligonal*, $C(x)$, continua, con pendiente en cada tramo igual al coste unitario de transporte. En consecuencia, si cada combinación de origen y destino tiene una función semejante, la función objetivo sería:

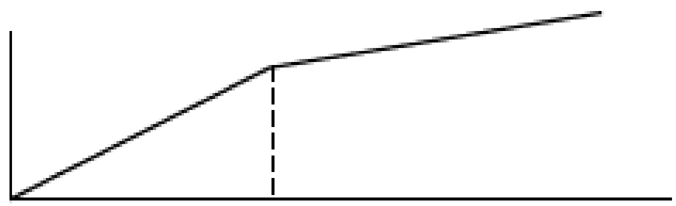
$$f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij}(x_{ij})$$

Introducción

Problema de transporte con descuentos por cantidad (II)

Hay que distinguir dos casos al hacer la formulación, según el descuento se aplique a todas las unidades si la cantidad supera un valor, o sólo a las que superan ese valor.

En el caso de que el descuento sólo se aplique a las cantidades que superan un valor, por ejemplo *k*, de modo que las *k* primeras siempre tienen un coste unitario *c* y las que sobrepasen a *k*, tengan un descuento siendo su coste unitario *c*-*a*, obsérvese que la función es continua:

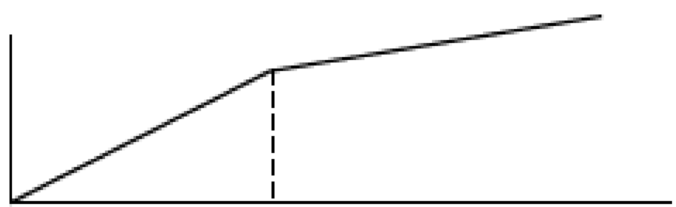


Introducción

Problema de transporte con descuentos por cantidad (III)

El modelo para representar esta función, siendo *X* la variable, pasa por dividir esta variable en dos, una hasta el valor *k* y otra para el exceso, obligando a que la segunda sea 0 mientras la otra llega al valor *k*:

$$\begin{aligned} & \min\ cX_1 + (c-a)X_2 \\ \text{s.a.}\quad & \mathrm{X} = X_1 + X_2 \\ & \delta k \leq X_1 \leq k \\ & 0 \leq X_2 \leq cot\ a\ \delta \\ & \delta \in \{0, 1\} \end{aligned}$$

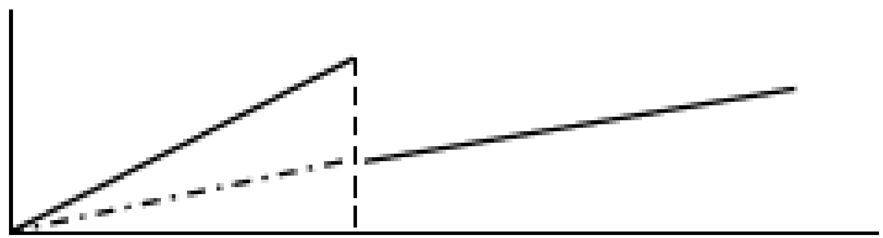


Introducción

Problema de transporte con descuentos por cantidad (IV)

En el caso de aplicarse el descuento a todas las unidades la función es discontinua, y el modelo es distinto, pues las variables no pueden ser cero a la vez:

$$\begin{gathered} min\ cX_1 + (c-a)X_2 \\ \text{s.a.} \quad X = X_1 + X_2 \\ 0 \leq X_1 \leq k\delta \\ k(1-\delta) \leq X_2 \leq cot\ a(1-\delta) \\ \delta \in \{0,1\} \end{gathered}$$



Introducción

Formulación general

$$\begin{aligned} &minimizar\ f(x) \\ &sujeto\ a\ x \in X, \end{aligned}$$

donde f es la función objetivo de n variables y X es un subconjunto de $\mathbb{R}^n$ que contiene las restricciones de las variables.

Introducción

Formulación general

$$\begin{aligned} &minimizar\ f(x) \\ &sujeto\ a\ x \in X, \end{aligned}$$

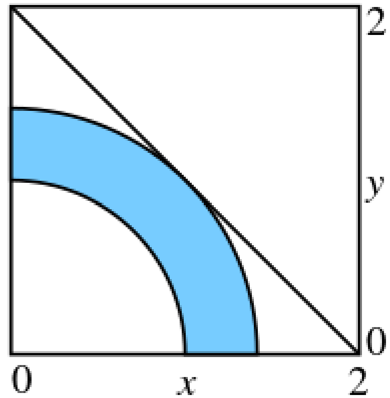
donde f es la función objetivo de n variables y X es un subconjunto de $\mathbb{R}^n$ que contiene las restricciones de las variables.

$$\begin{aligned} &minimizar\ f(x) \\ &s.a.\ g_i(x) \leq b_i, i = 1, \dots, m \end{aligned}$$

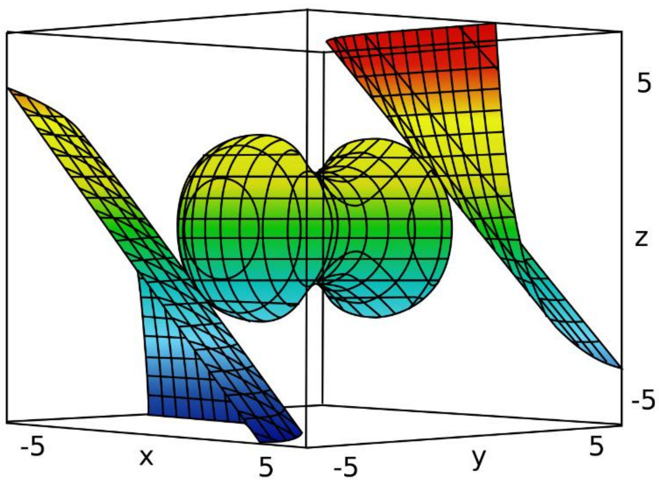
donde la función objetivo y/o alguna restricción es no lineal.

Ejemplos

$$
\begin{aligned}
&\min f(x) = x_1 + x_2\\
&s.t.\ x_1^2 + x_2^2 \geq 1\\
&\qquad x_1^2 + x_2^2 \leq 2\\
&\qquad x_1 \geq 0, x_2 \geq 0
\end{aligned}
$$



$$
\begin{aligned}
&\min f(x) = x_1x_2 + x_2x_3\\
&s.t.\ x_1^2 - x_2^2 + x_3^2 \leq 2\\
&\qquad x_1^2 + x_2^2 + x_3^2 \leq 10
\end{aligned}
$$



2 Definiciones clásicas

Continuidad

Una función f es continua en un punto x_0 de su dominio D si

$$\forall \epsilon > 0 \quad \exists \delta > 0 \quad t.q. |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon$$

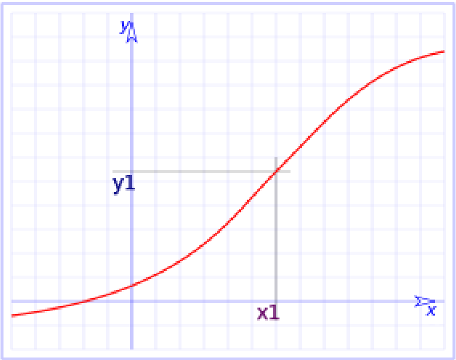
Intuitivamente: conforme te acercas a x_0 el valor de f se aproxima a $f(x_0)$:

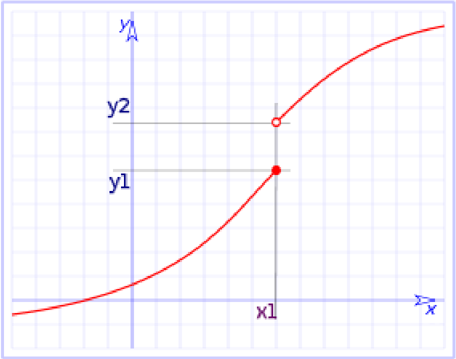
$$\lim_{x \to x_0} f(x) = f(x_0)$$

Diremos que es continua en la región factible X si:

- Es continua para todo punto en el interior de X.
- Lo es también para todo punto en el borde de X (existe el límite conforme te acercas a él desde el interior de X).

Informalmente: <u>no hay “huecos” ni “saltos” dentro de X</u>.





Linealidad

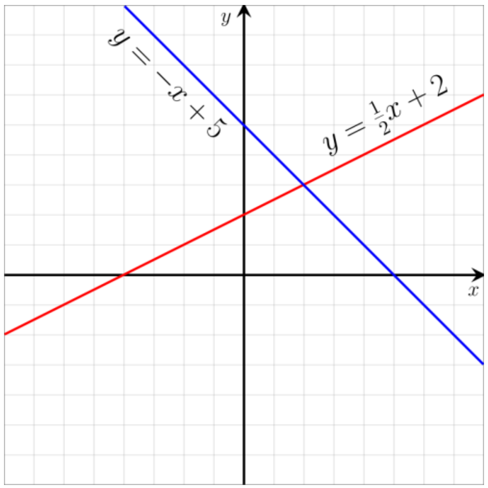
Una función f de una variable x es lineal con respecto a x si es de la forma:

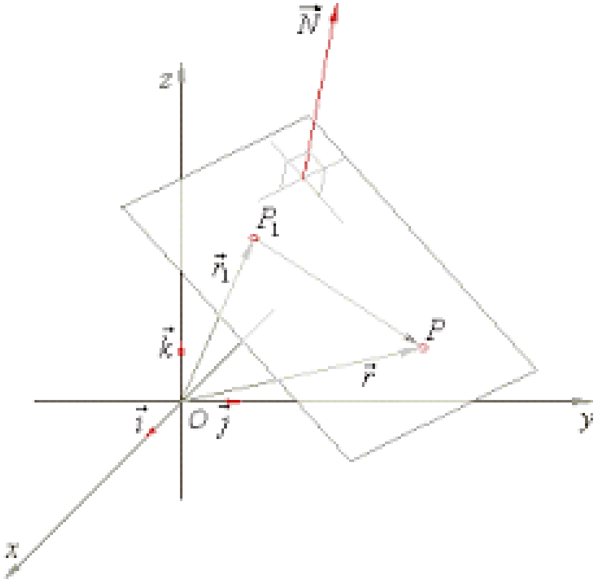
$$f(x) = ax + b$$

Con a la pendiente de la recta correspondiente y b el intercepto/bias. Para varias variables:

$$f(x_1, \dots, x_n) = a_1 x_1 + \ \dots + a_n x_n + b = \boldsymbol{a^T x} + b$$

Donde el vector $a^T = (a_1, \dots, a_n)$ es el vector perpendicular al hiperplano de dimensión n que representa a f.





Cuadráticas

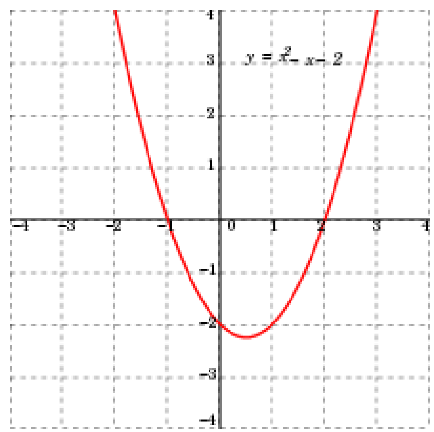
Una función f de una variable x es cuadrática con respecto a x si es de la forma:

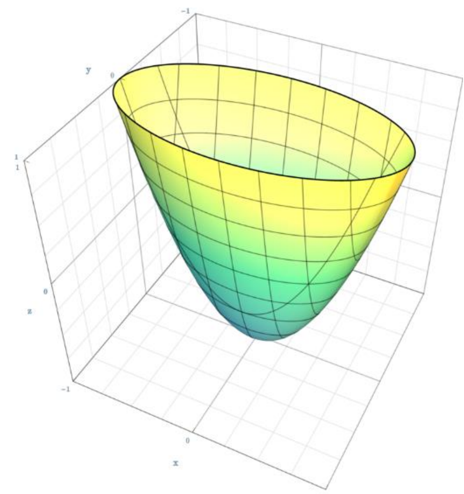
$$f(x) = ax^2 + bx + c, \qquad a \neq 0$$

Donde a, b, c determinan la forma y posición de la parábola correspondiente. Para varias variables:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{p}^{\boldsymbol{T}}\boldsymbol{x} + R$$

Donde Q es una matriz n x n, p un vector n x 1, y R una constante que determinan la forma cuádrica correspondiente.





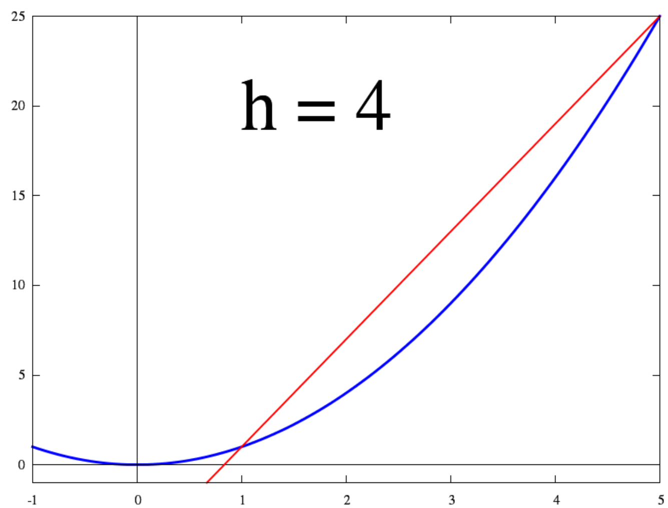
Diferenciabilidad (una variable)

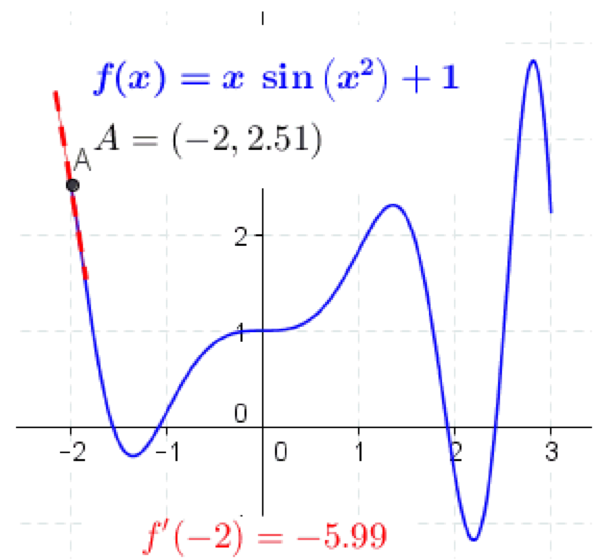
Una función de una variable f(x) es derivable en un punto x_0 de su dominio si existe el límite:

$$L = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

De existir, L es la derivada de f en x_0, escrita habitualmente como $f'(x_0)$ o como $\frac{\partial f}{\partial x}(x_0)$

Intuitivamente: la derivada es la pendiente de la recta tangente a f en x_0. Existe si la función cambia sus valores "suavemente".





Diferenciabilidad (varias variables)

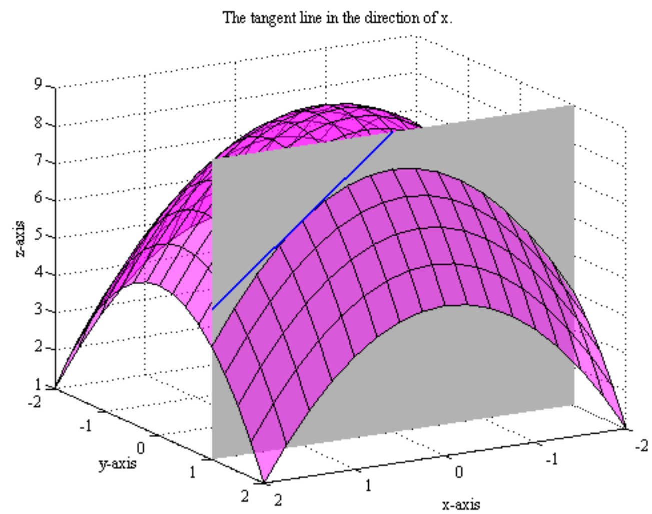
Una función de varias variables $f(x_1, \ldots, x_n)$ es derivable en $\boldsymbol{x_0}$ si existe el límite vectorial:

$$L = \lim_{\boldsymbol{h} \to 0} \frac{f(\boldsymbol{x_0} + \boldsymbol{h}) - f(\boldsymbol{x_0})}{\boldsymbol{h}}$$

En ese caso, decimos que el gradiente de f en x_0, escrito habitualmente como $\nabla f(x_0)$, es el vector de derivadas parciales:

$$\nabla f(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0) \right)$$

Intuitivamente: <u>la derivada parcial es la pendiente del plano tangente a la superficie de f en la dirección indicada por ella</u> (cuando sólo cambia la variable x_i). <u>El gradiente condensa cuánto cambia la función si uno se desplaza en las n direcciones determinadas por las variables</u>.



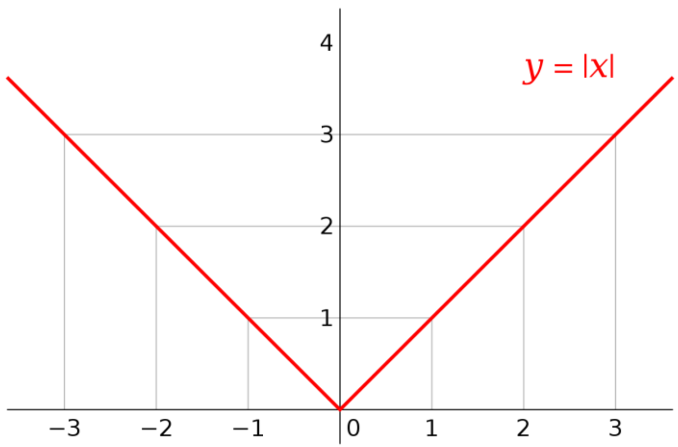
Diferenciabilidad y continuidad

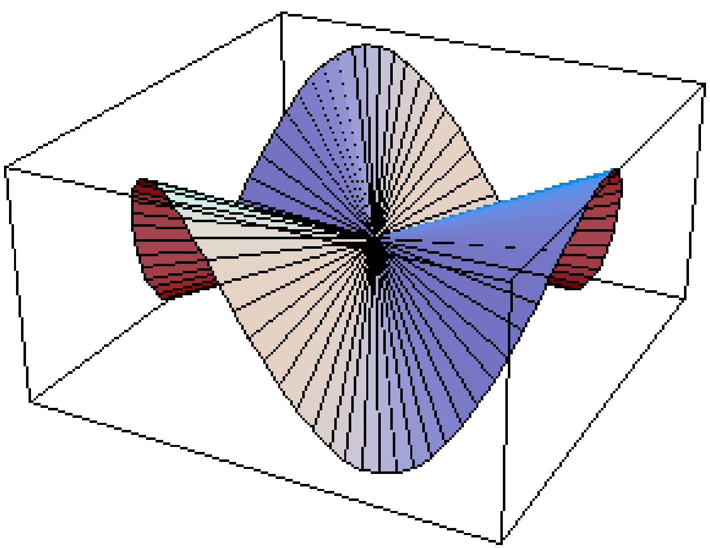
Si una función es diferenciable en un punto, entonces tiene que ser continua en ese punto.

Por lo tanto, si una función es diferenciable en una determinada región, también tiene que ser continua en ese región.

Lo contrario NO es cierto: una función puede ser continua, pero no diferenciable.

Intuitivamente: la función puede no tener “saltos”, pero sí “picos” en donde no cambia de forma “suave”.





Doble diferenciabilidad

Una función es doblemente diferenciable si existe la derivada de su derivada f''(x) para el caso de una variable, o el gradiente de su gradiente en el caso de varias. En este último caso, obtenemos la matriz hessiana (o hessiano), que se forma con todos los posibles pares de derivadas parciales:

$$H_f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1 \partial x_1}(x_0) & \cdots & \frac{\partial f}{\partial x_1 \partial x_n}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_n \partial x_1}(x_0) & \cdots & \frac{\partial f}{\partial x_n \partial x_n}(x_0) \end{bmatrix}$$

<u>Si f es doblemente diferenciable en un punto, entonces es que la 1ª derivada es continua</u>.

Intuitivamente: nos da información sobre la curvatura concreta en el punto, y si existe es porque esa curvatura cambia "suavemente" (más sobre esto pronto).

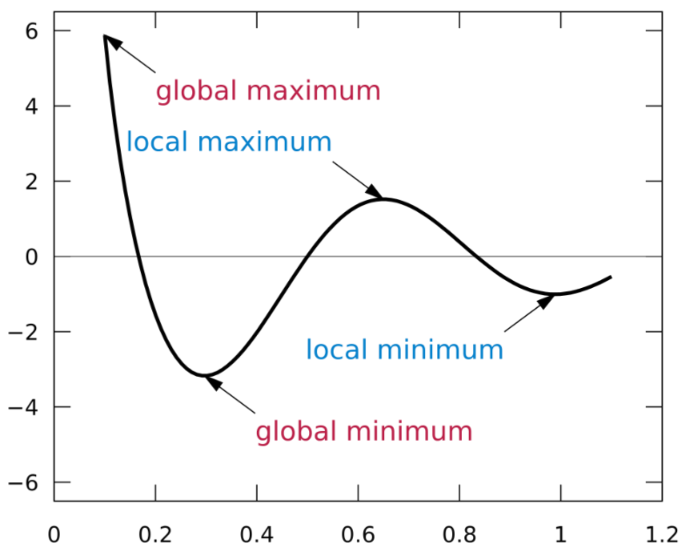
Óptimos locales y globales (extremos)

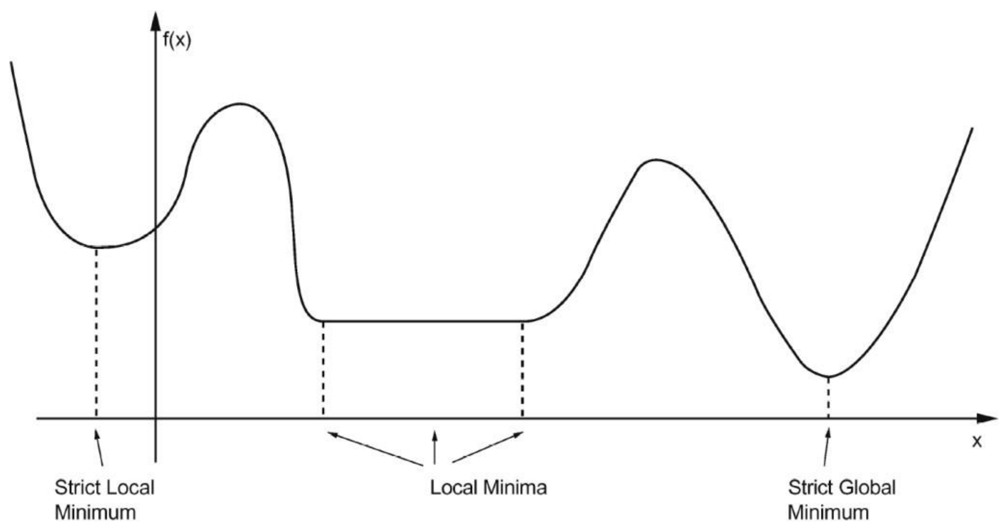
Un punto x^* es un mínimo local de f si no hay puntos menores en su entorno, es decir,

$$\text{existe } \epsilon > 0 \text{ } tal \text{ } que \text{ } f(x^*) \leq f(x) \text{ } para \text{ } todo \text{ } x \in X \text{ } con \text{ } \left|\left|x - x^*\right|\right| < \epsilon$$

Un punto x^* es un mínimo global de f si no hay puntos mejores en toda la región factible, es decir,

$$f(x^*) \leq f(x) \text{ } para \text{ } todo \text{ } x \in X$$





Convexidad y concavidad

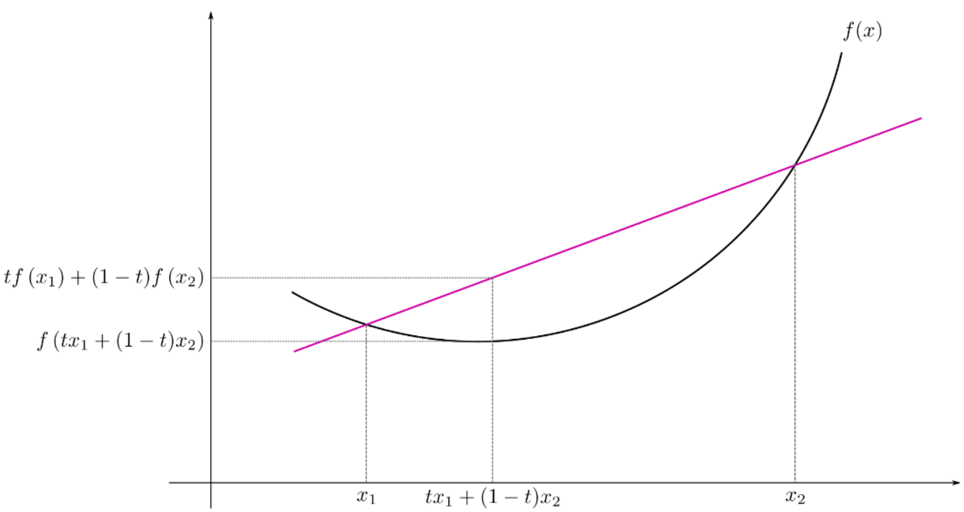
Una función es convexa si cumple que, para cualesquiera 2 puntos de su dominio:

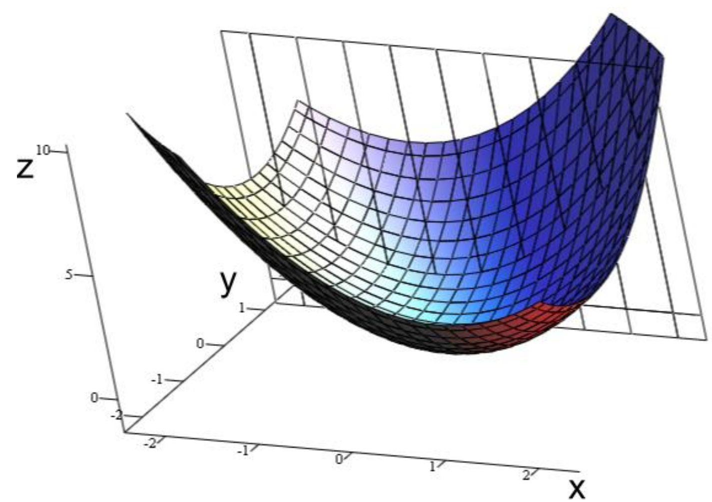
$$f(t\boldsymbol{x_1} + (1-t)\boldsymbol{x_2}) \leq tf(\boldsymbol{x_1}) + (1-t)f(\boldsymbol{x_2}) \qquad \forall t \in [0,1]$$

Si fuera al revés (con >=), hablamos de función cóncava.

Intuitivamente: <u>si unimos por una línea recta cualesquiera 2 puntos de la función, la superficie de f siempre "cae por debajo" de esa línea (convexidad) o "por encima" (concavidad)</u>.

<u>En una función convexa/cóncava, todo mínimo/máximo local tiene que ser también global</u>.



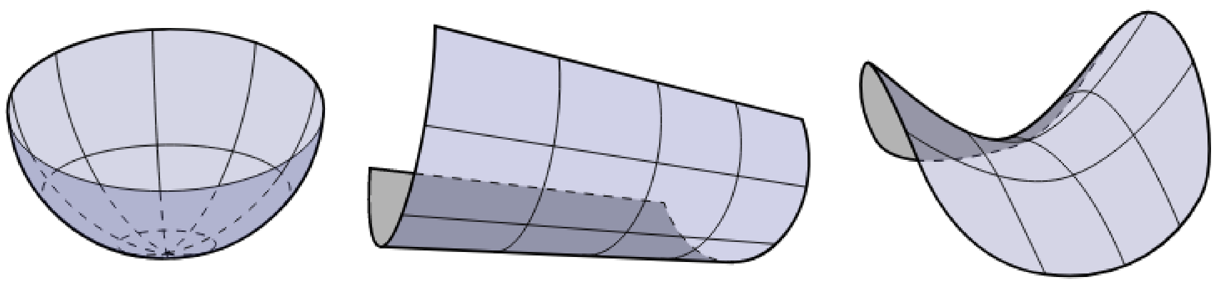


Definición de una matriz

Una matriz simétrica Q es:

- Definida positiva si $\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} > 0 < \forall x \neq \mathbf{0}$
- Semidefinida positiva si $\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} \geq 0 < \forall x$
- Definida negativa si $\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} < 0 < \forall x \neq \mathbf{0}$
- Semidefinida negativa si $\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} \leq 0 < \forall x$
- Indefinida si no se cumple ninguna de las anteriores

Intuición: la definición de la matriz Q de una cuadrática, o más en general del hessiano de la función objetivo, nos dará pistas sobre la convexidad/concavidad en el entorno del punto, y por tanto saber si estamos cerca de un máximo, de un mínimo o de algún otro tipo de punto.



$x^2 + y^2$ (definite) — x^2 (semidefinite) — $x^2 - y^2$ (indefinite)

Tipos de optimización (I)

$$minimizar\ f(x)$$
$$sujeto\ a\ x \in X$$

- Si f es lineal y las restricciones $x \in X$ son lineales, estamos frente a un problema de optimización lineal (LP). Tenemos que:
  - f es lineal, luego también es continua (y convexa).
  - las restricciones igual, y sus bordes forman un poliedro X (si el problema es factible y acotado).
  - hay un óptimo global (aunque puede que no estricto) en un vértice del poliedro.

- Si a lo anterior añadimos que alguna de las variables x es entera, tenemos un problema de optimización entera (IP) (también llamada discreta). Tenemos que:
  - Allá donde influyan las variables enteras (función objetivo f y/o restricciones) se pierde la continuidad y la convexidad, lo cual invalida todo el razonamiento de que hay un óptimo global en un vértice (de hecho, la región factible ya no es tampoco un poliedro, y ya un óptimo local no tiene tampoco por qué ser global).

Tipos de optimización (II)

$$minimizar\ f(x)$$
$$sujeto\ a\ x \in X$$

- Si f es convexa, y las restricciones $x \in X$ se dividen en convexas de desigualdad (no necesariamente lineales) y lineales de igualdad (estas sí), hablamos de optimización convexa.

- Si f es cuadrática y las restricciones $x \in X$ son lineales, estamos frente a un problema de optimización cuadrática (QP). Tenemos que:
  - Si la matriz Q que define la cuadrática de f es (semi)definida positiva, entonces f es (estrictamente) convexa. En ese caso, como las restricciones siguen siendo lineales (y también convexas), cualquier óptimo local es global.
  - Si es definida negativa o indefinida, entonces f no es convexa.

- En otro caso, hablamos en general de optimización no lineal (NLP), dividiendo en:
  - No restringida si no hay restricciones $x \in X$
  - Restringida si sí que las hay.

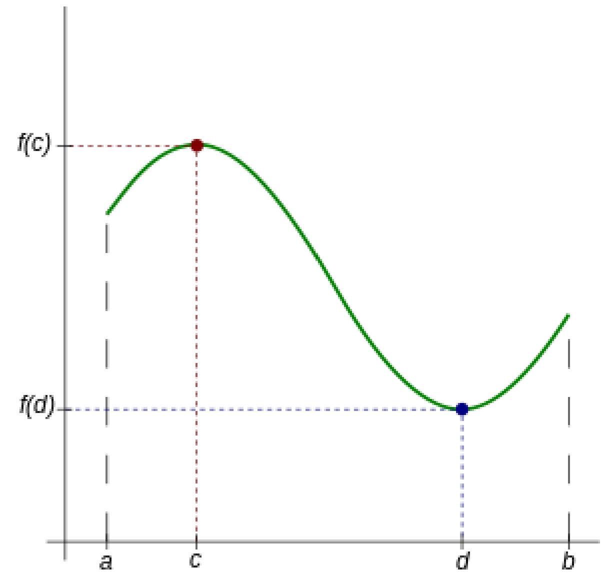
- En NLP en general sólo podemos aspirar a obtener soluciones locales, pero rara vez globales.

3 Búsqueda de extremos

Teoremas

Teorema de Weierstrass (valores extremos): si f es continua en un intervalo cerrado y acotado [a, b], entonces hay al menos dos puntos x_1 y x_2 donde f alcanza extremos absolutos, es decir,

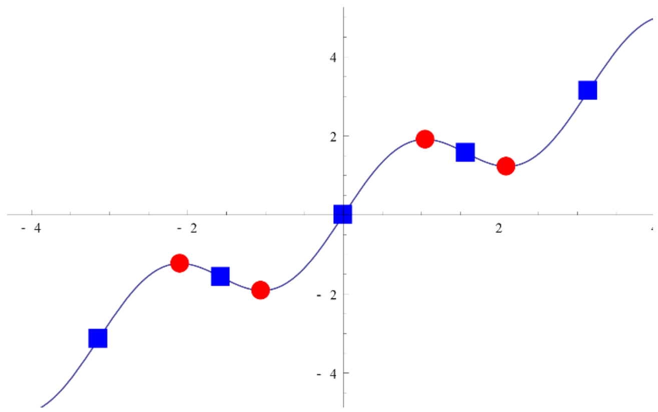
$$f(x_1) \leq f(x) \leq f(x_2) \;\; \forall x \in [a, b]$$



Esto nos dice que si la región factible X está acotada, el mínimo/máximo debe existir y al menos debe haber uno, pero NO nos dice dónde está ni qué condiciones debe cumplir.

Teorema de Fermat (extremos interiores): si f tiene un máximo/mínimo en un punto y es derivable en él, entonces la derivada es 0. Por lo tanto los extremos de f los podemos tener en:

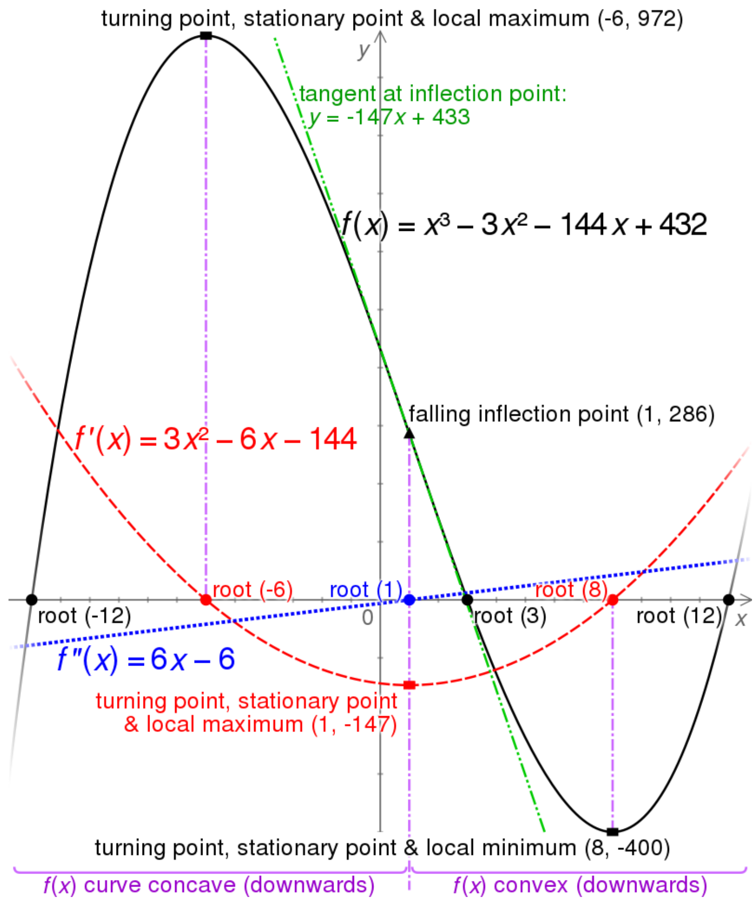
- El borde de la región factible X
- Los extremos interiores de X
- En puntos donde X no sea derivable



Puntos estacionarios (1 variable)

Supongamos f diferenciable. Una 1ª idea entonces es calcular todos aquellos puntos para los que f'(x) = 0 (puntos estacionarios) y ver lo que pasa en ellos:

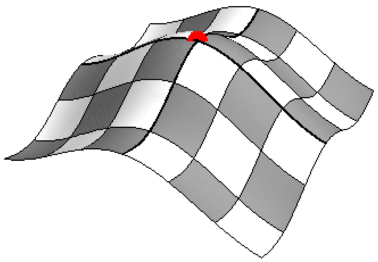
- En f(x) cambia la curvatura de la función:
    - En un máximo la derivada pasa de ser positiva a negativa (de crecer a decrecer).
    - En un mínimo de ser negativa a positiva (de decrecer a crecer).
    - En un punto de inflexión de ser cóncava a convexa, o viceversa.
- Si existe la 2ª derivada f''(x):
    - Si f''(x) < 0, estamos en un máximo local.
    - Si f''(x) > 0, en un mínimo local.
    - Si f''(x) = 0, podría ser un máximo, un mínimo o un punto de inflexión.

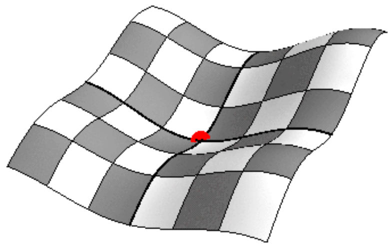


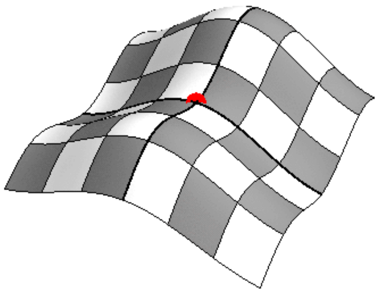
Puntos estacionarios (varias variables)

Aquí tendríamos que ver para qué puntos todas las derivadas parciales son 0, es decir, $\nabla f(x) = \mathbf{0}$:

- En f(x) cambia la curvatura de la función:
  - En un máximo da igual en qué dirección vayamos, que la función decrecerá.
  - En un mínimo da igual en qué dirección vayamos, que la función crecerá.
  - En un punto de silla dependiendo de en qué dirección vayamos, decrecerá o crecerá.
- Si existe el hessiano $H_f(x)$:
  - Si $H_f(x)$ es definido negativo, es un mínimo local.
  - Si $H_f(x)$ es definido positivo, es un mínimo local.
  - Si tiene autovalores positivos y negativos (indefinido) es un punto de inflexión (punto de silla).
  - En otro caso, podría ser cualquiera de ellos.







Más derivadas (1 variable)

El procedimiento anterior se puede generalizar si la función es n+1 veces derivable y tenemos que:

- $f'(x) = \ldots = f^{(n)}(x) = 0$ (todas las derivadas hasta la n-ésima se hacen 0)
- $f^{(n+1)}(x)$ distinto de 0 (la n+1 ya no lo es)
- Entonces:
    - Si n es impar y $f^{(n+1)}(x) < 0$, estamos en un máximo local.
    - Si n es impar y $f^{(n+1)}(x) > 0$, en un mínimo local.
    - Si n es par y $f^{(n+1)}(x) < 0$, en un punto de inflexión en el que la función decrece.
    - Si n es par y $f^{(n+1)}(x) > 0$, en un punto de inflexión en el que la función crece.
- Problemas con esta aproximación:
    - En varias variables implica tensores (gradientes de gradientes de gradientes…)
    - Encontrar los puntos en que las derivadas se hacen 0 es complejo (suele implicar resolver sistemas de ecuaciones).
    - No vale para optimización con restricciones (los puntos estacionarios pueden no ser factibles).
    - Algunas funciones pueden no ser diferenciables n veces.

Multiplicadores de Lagrange

Es un procedimiento para convertir un problema con restricciones de igualdad en uno sin restricciones. Suponiendo:

$$\begin{aligned} & \max f(x) \\ s.t. \quad & g(x) = 0 \end{aligned}$$

Podemos crearnos el Lagrangiano $L(x) = f(x) - \lambda g(x)$ y buscar sus puntos estacionarios, es decir, donde $L'(x) = f'(x) - \lambda g'(x) = 0$. En concreto, el óptimo estará en un punto de silla.

Generalizando a n variables y m restricciones, si tenemos un problema:

$$\begin{aligned} & \max f(\boldsymbol{x}) \\ s.t. \quad & g_i(\boldsymbol{x}) = 0 \quad \forall i \in 1, \dots, m \end{aligned}$$

El Lagrangiano sería $L(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathrm{f}(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x})$, y habría que encontrar los puntos donde (sistema de n+m ecuaciones con n+m incógnitas):

$$\nabla L(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i \nabla g_i(\boldsymbol{x}) = \mathbf{0}$$

Condiciones de Karush-Kuhn-Tucker (KKT)

Lo anterior se puede generalizar todavía más a problemas con restricciones de igualdad y desigualdad. Suponiendo:

$$\begin{aligned} & \min f(\boldsymbol{x}) \\ s.t. \quad & g_i(\boldsymbol{x}) \leq 0 \quad \forall i \in 1, \dots, m \\ & h_j(\boldsymbol{x}) = 0 \quad \forall j \in 1, \dots, l \end{aligned}$$

El Lagrangiano sería $L(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathrm{f}(\mathbf{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x})$, y habría que encontrar los puntos donde se cumple (sistema de n+m+l ecuaciones con n+m+l incógnitas):

$$\nabla L(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\boldsymbol{x}) = \mathbf{0} \quad (\boldsymbol{estacionaridad})$$

$$g_i(\boldsymbol{x}) \leq \mathbf{0} \quad \forall i \in 1, \dots, m, \quad h_j(\boldsymbol{x}) = 0 \quad \forall j \in 1, \dots, l \quad (\boldsymbol{factibilidad\ primal})$$

$$\mu_i(\boldsymbol{x}) \geq \mathbf{0} \quad \forall i \in 1, \dots, m, \quad (\boldsymbol{factibilidad\ dual})$$

$$\mu_i g_i(\boldsymbol{x}) = 0 \quad \forall i \in 1, \dots, m, \quad (\textbf{laxitud complementaria})$$

Sólo hacen falta primeras derivadas y ya hemos cubierto el caso con restricciones, pero esto sigue implicando resolver un sistema de ecuaciones complejo…