

OPTIMIZACIÓN PARA INGENIEROS

(Notas de clase)

Instructores:
Luis Zerpa
Juan Colmenares

Enero 2004

Índice General

1. OPTIMIZACIÓN.....	3
2. Mathematical Optimization Problem or Mathematical Program.....	4
2.1. Formulación General de un Problema de Optimización.....	4
2.2. Programación Nolineal (<i>Non Linear Programming</i> NLP).....	5
2.2.1. Características de los problemas que trataremos mayormente en el curso.....	5
2.2.2. Tipos de Problemas Nolineales.....	6
2.2.3. Algoritmos Iterativos y Convergencia.....	8
3. Bibliografía.....	9
4. Funciones de una variable.....	10
4.1. Continuidad de una función en un número.....	10
4.2. Teoremas sobre Continuidad.....	11
4.3. Continuidad en un Intervalo.....	11
4.4. Diferenciabilidad y Continuidad.....	12
5. Derivada de una Función.....	13
6. Valores Máximos y Mínimos de una Función de una Variable.....	13
6.1. Extremos Relativos.....	13
6.2. Extremos Absolutos.....	16
6.2.1. Extremos Absolutos en un Intervalo.....	16
6.2.2. Procedimientos para determinación de extremos absolutos en intervalo cerrado:.....	16
7. Funciones Crecientes y Decrecientes y Criterio de la Primera Derivada.....	18
7.1. Criterio de la Primera derivada para Extremos Relativos.....	18
7.1.1. Procedimiento para Determinar Extremos Relativos.....	19
7.2. Criterio de la Segunda Derivada.....	19
8. Formula de Taylor (Brook Taylor 1685 – 1731).....	19
8.1. Polinomio de Taylor.....	20
9. Funciones de Varias Variables (Campos Escalares) $R^n \rightarrow R^1$	20
9.1. Continuidad de Campos Escalares.....	20
9.2. Operaciones sobre Funciones Continuas.....	20

9.3.	Gradiente.....	21
9.4.	Formula de Taylor en Varias Variables.....	22
9.5.	Extensión de los Criterios de Existencia de Máximo y Mínimos.....	23
10.	Formas Cuadráticas.....	24
10.1.	Propiedades de las formas cuadráticas.....	25
10.2.	Métodos de clasificación de formas cuadráticas.....	26
10.2.1.	Método de los menores principales.....	26
10.2.2.	Método de los autovalores:.....	26
11.	Funciones Convexas.....	27
11.1.	Caracterización alternativa de las funciones convexas.....	28
11.1.1.	Criterios de la primera y segunda derivada.....	29

1. OPTIMIZACIÓN

La mayoría de los problemas en el mundo real tienen varias soluciones y algunos tienen infinitas soluciones. El propósito de la optimización es encontrar o identificar la mejor solución posible, entre todas las soluciones potenciales, para un problema dado, en términos de algún o algunos criterios de efectividad o desempeño.

Existen numerosas estrategias de optimización que van desde sofisticados procedimientos matemáticos (tanto analíticos como numéricos) hasta la simple pero inteligente aplicación de la aritmética.

Asumiendo que un problema de optimización está definido de alguna manera, una clasificación general de los métodos de optimización es la siguiente:

Métodos Analíticos: Uso del cálculo diferencial
(insuficiente para problemas no lineales).

Métodos Numéricos: Algoritmos
(Procedimientos iterativos)
(Sobre esto se centra nuestro curso).

Otros: Métodos gráficos, métodos experimentales, estudio de casos.

Casi siempre nuestro interés en la optimización se centra en la solución de problemas reales, los cuáles deben ser representados matemáticamente. La optimización de la representación matemática de procesos reales presenta 2 tipos de dificultades:

1. Formulación del modelo matemático (representatividad)
(Función a ser optimizada ó función objetivo)

2. Técnica de Solución:

- Existencia de varios extremos locales y globales.
- Se suponen que los coeficientes y variables del modelo (función objetivo) no son variables aleatorias.
- Errores de redondeo de la aritmética punto flotante.

2. Mathematical Optimization Problem or Mathematical Program

Maximización o minimización de funciones algebraicas de una o más variables. La selección de las variables estarán restringidas por ecuaciones o inecuaciones algebraicas llamadas restricciones, de forma tal que el objetivo no es encontrar el mejor valor posible sino el mejor valor permitido por las restricciones.

2.1. Formulación General de un Problema de Optimización

Encontrar (c_1, \dots, c_n) tal que

se maximice o minimice $f(c_1, \dots, c_n)$

sujeto a restricciones:

$$g_i(c_1, \dots, c_n) \begin{cases} = \\ \geq \\ \leq \end{cases} b_i \quad \begin{array}{l} \text{donde } i = 1, \dots, m \\ b_i \text{ es escalar} \\ \chi_i, f \text{ y } g_i \text{ son reales} \\ S \text{ dominio de } f \end{array}$$

en forma vectorial,

Encontrar \mathbf{x} tal que

Max ó min $f(\mathbf{x})$

$$\text{Sujeto a } g_i(\mathbf{x}) \begin{cases} = \\ \leq \\ \geq \end{cases} b_i \quad \begin{array}{l} \text{donde} \\ \mathbf{x} = (x_1, \dots, x_n) \\ \mathbf{x} \in S \end{array}$$

La rama de la matemática aplicada que resuelve este tipo de problemas se denomina “Programación Matemática”(Mathematical Programming).

$f(\mathbf{x})$ es referida como *función objetivo*, una solución posible al problema antes formulado se denomina *solución óptima* u *óptimo*.

La *optimización computacional* es una herramienta fundamental en los *procesos de toma de decisiones*.

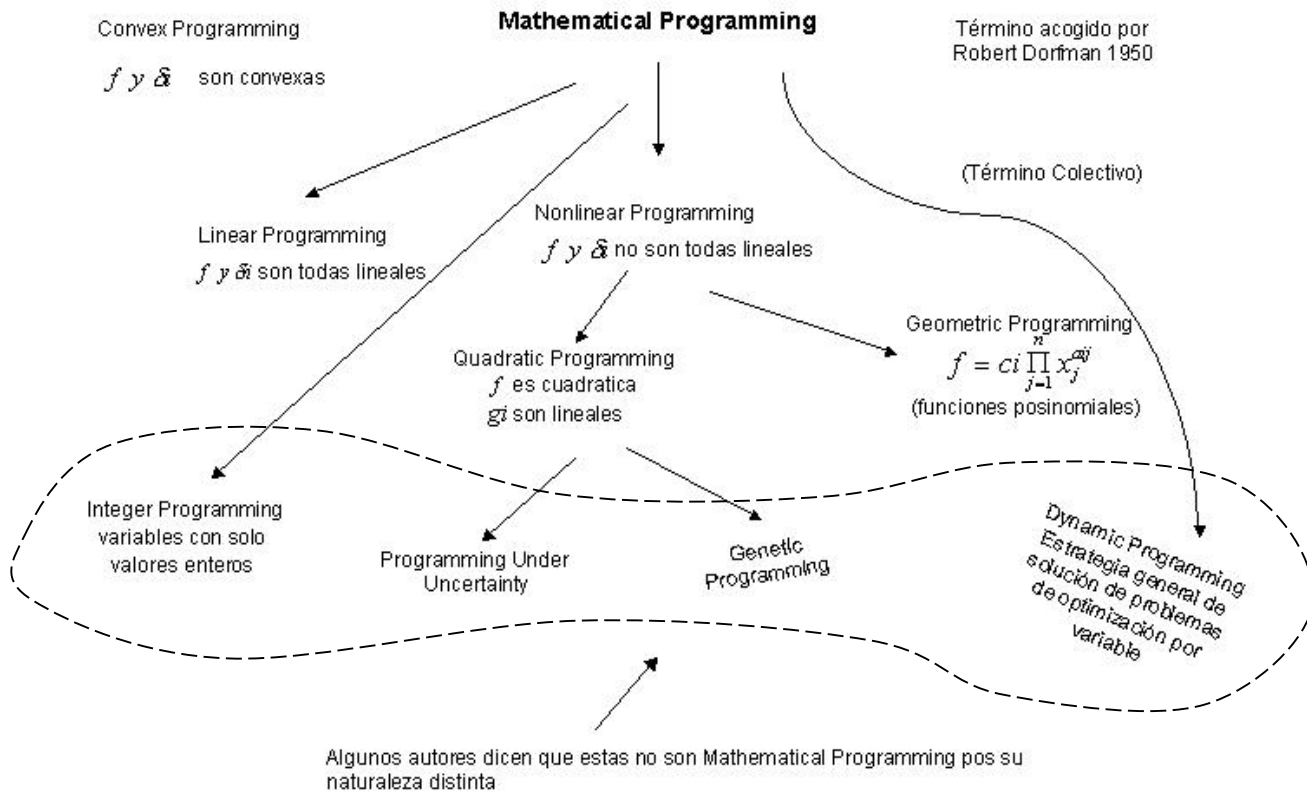


Figura 1 – Tipos de Programación Matemática

2.2. Programación No lineal (*Non Linear Programming* NLP)

NLP: Conjunto de técnicas para optimizar funciones no lineales sujetas a restricciones de igualdad o desigualdad. Tanto las funciones como las restricciones pueden ser de una o más variables.

El curso trata fundamentalmente de *programación no lineal*.

2.2.1. Características de los problemas que trataremos mayormente en el curso

- Funciones objetivo y restricciones continuas con sus primeras derivadas parciales también continuas (suaves)¹
- Inecuaciones estrictas no son permitidas (< ó >) solo se permiten restricciones \geq , \leq e $=$.
- El problema debe ser determinístico.

¹ Esto garantiza que pequeños cambios en \underline{x} conlleve a pequeños cambios en valores asociados

- Todas las variables deben ser reales, ninguna puede tomar únicamente valores enteros. (*Continuous Programming*).
- S dominio de f y g_i sea una región conectada.

2.2.2. Tipos de Problemas Nolineales

Sin restricciones y con restricciones

Tamaño de los Problemas

Una forma de medir la complejidad de los problemas es en función del número de variables o del número de restricciones.

Pequeña Escala: 5 variables y restricciones.

Escala intermedia: de 5 a 100 variables y restricciones.

Gran Escala: más de 100 y quizás 1000 variables y restricciones.

Esta clasificación no es rígida pero refleja las diferencias de los enfoques de solución para cada tipo de problema.

Pequeña Escala: Resuelto a mano

Mediana Escala: Computador Personal o Servidor de Propósito General

Gran Escala: Mainframe para cálculo científico (cray) y se explota la estructura del problema. Algoritmos paralelos.

En el curso se estudiarán la teoría y los métodos que permiten efectivamente la solución de la más amplia variedad de problemas (pequeña y mediana escala principalmente). A pesar de que un gran número de algoritmos han sido propuestos para la solución del problema general de optimización no lineal, sólo unos pocos han demostrado ser efectivos cuando se aplican a problemas de “gran – escala”.

No existe un método general de optimización no lineal en el sentido como es SIMPLEX para problemas lineales. (Se tienen más de 50 años trabajando en la solución de problemas lineales).

Ninguno es tan superior para ser clasificado como la panacea universal de la NLP.

Esto hace que sea un amplio y fructífero campo de investigación, el cual avanza según el siguiente algoritmo:

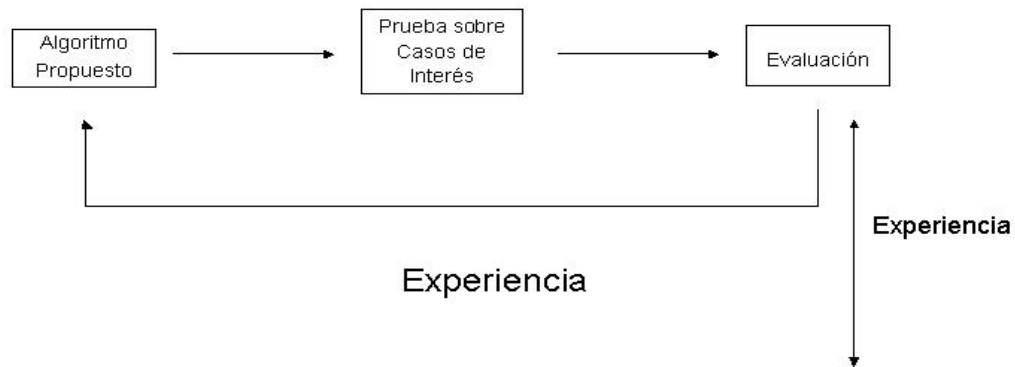


Figura 2 – Avance de la Programación no lineal

En NLP:

Si una estrategia no presenta prueba de convergencia, esta puede ser efectiva, e.g. EGO y DIRECT.

Más aún la garantía de convergencia de un algoritmo NLP para un caso especial no ofrece, en muchos casos, mayor entendimiento con relación a la creación de estrategias efectivas para problemas más complejos.

Criterios de Comparación de Algoritmos

1. Número de evaluaciones de la función objetivo
2. Confiabilidad (Éxito en alcanzar la solución)
3. Rapidez
4. Tiempo de Preparación del usuario (sobre parametrización)
5. Precisión de la solución
6. Grado de satisfacción de las restricciones
7. Dificultad

2.2.3. Algoritmos Iterativos y Convergencia

La mayoría de los algoritmos de NLP son iterativos.

$$\vec{x}_1 \rightarrow \vec{x}_2 \rightarrow \mathbf{K} \rightarrow \vec{x}_n \rightarrow \vec{x}^*$$

Solución Óptima

En programación lineal existe una secuencia de longitud finita para alcanzar la solución.

En NLP la secuencia generalmente no alcanza la solución óptima sino que converge hacia ella.

En problemas no lineales se determina una solución lo suficientemente cercana a la óptima.

La teoría de algoritmos iterativos se divide en:

1. Diseño del Algoritmo
2. Convergencia Global: Análisis de convergencia global (si eventualmente converge)
3. Convergencia Local: Análisis de convergencia local (la razón a la cual el algoritmo converge en la Solución óptima).

“Una buena teoría es mejor que miles de corridas”

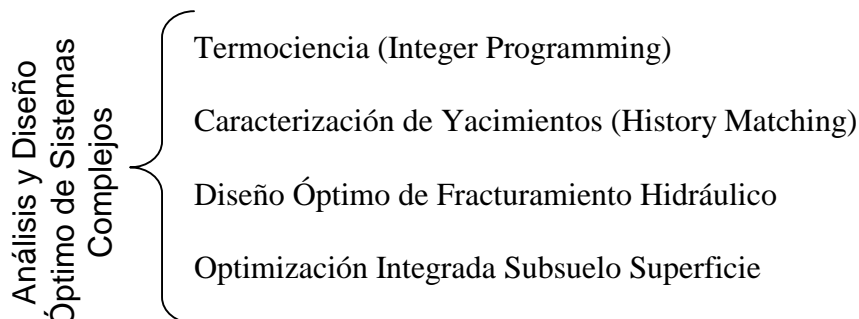
Esto da una idea de la tratabilidad de los problemas mediante un análisis simple lo cual es muy importante.

Un enfoque experimental también es posible, pero no ha habido suficiente tiempo para recopilar tal evidencia debido a la aparición frecuente de nuevos y mejores Algoritmos de Programación Nolineal.

Áreas de Aplicación de Programación Nolineal

- Ingeniería
- Programación de producción, Pronóstico, Control de Inventario, Control de Calidad, Control de Procesos, Mantenimiento, Diseño, Estimación de Presupuesto.

Optimización en el ICA



3. Bibliografía

- ⊖ David Luenberger. *Linear and nonlinear Programming*. 2nd Ed. Addison-Wesley
- ⊖ Donald Simmons. *Nonlinear Programming for Operations Research*. Prentice-Mall.
- ⊖ David Himmelblan. *Applied Nonlinear Programming*. McGraw-Hill.
- ⊖ Panos Papalambros y Douglas Wilde. *Principles of Optimal Design*. Cambridge.

4. Funciones de una variable

4.1. Continuidad de una función en un número

Se dice que f es continua en el número a si y solo si las siguientes 3 condiciones se satisfacen:

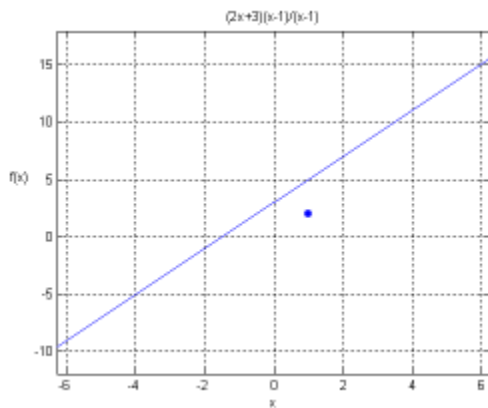
i. $f(a)$ existe

ii. $\lim_{x \rightarrow a} f(x)$ existe

iii. $\lim_{x \rightarrow a} f(x) = f(a)$

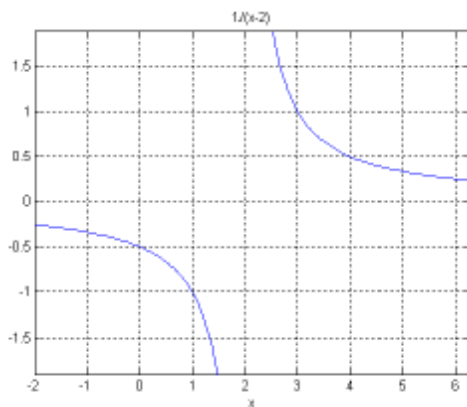
Si al menos una de estas 3 condiciones no se cumple se dice que f es discontinua en a .

Claramente, la noción geométrica de un *salto* en la gráfica de la función es sinónimo de discontinuidad, por ejemplo,



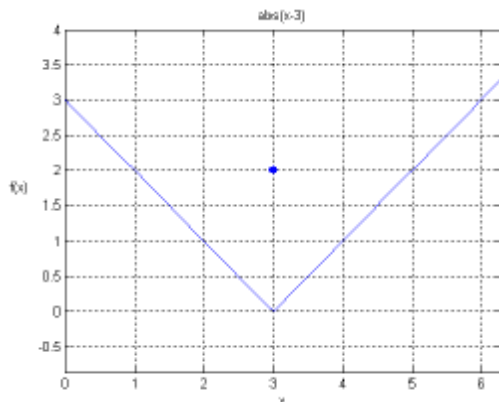
$$f(x) = \begin{cases} \frac{(2x+3)(x-1)}{(x-1)}; & \text{si } x \neq 1 \\ 2; & \text{si } x = 1 \end{cases}$$

Discontinuidad removible



$$f(x) = \frac{1}{x-2}$$

Discontinuidad Esencial



$$f(x) = \begin{cases} |x-3| & \text{si } x \neq 3 \\ 2 & \text{si } x = 3 \end{cases}$$

4.2. Teoremas sobre Continuidad

Teorema: Si f y g son continuas en a entonces:

- i) $f + g$ es continua en a
- ii) $f - g$ es continua en a
- iii) $f \times g$ es continua en a
- iv) $f \div g$ es continua en a suponiendo que $g(a) \neq 0$

Teorema: Una función polinomial es continua en todo número

Definición: Se dice que una función es continua en un número a , si f está definida en algún intervalo abierto que contiene en a y si para cualquier $\epsilon > 0$ existe un $\delta > 0$ tal que:

$$|f(x) - f(a)| < \epsilon \text{ siempre que } |x - a| < \delta$$

4.3. Continuidad en un Intervalo

Definición:

Continuidad por la derecha

Se dice que f es continua por la derecha del número a si y solo si satisface las siguientes condiciones:

- i) $f(a)$ existe
- ii) $\lim_{x \rightarrow a^+} f(x)$ existe
- iii) $\lim_{x \rightarrow a^+} f(x) = f(a)$

Continuidad por la izquierda

Se dice que f es continua por la izquierda del número a si y solo si,

i) $f(a)$ existe

ii) $\lim_{x \rightarrow a^-} f(x)$ existe

iii) $\lim_{x \rightarrow a^-} f(x) = f(a)$

Definición: Una función cuyo dominio incluye el intervalo cerrado $[a, b]$ se dice que es continua en $[a, b]$ si y solo si es continua en el intervalo abierto (a, b) , así como es continua por la derecha de a y continua por la izquierda de b .



Definición: f es continua en $[a, b)$ si y solo si es continua en (a, b) y continua por la derecha de a .

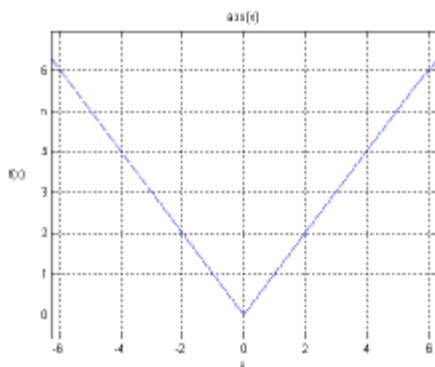
Definición: f es continua en $(a, b]$ si y solo si es continua en (a, b) y continua por la izquierda de b .

4.4. Diferenciabilidad y Continuidad

La continuidad de una función no implica la diferenciabilidad de dicha función en ese número.

Ej.

$$f(x) = |x|$$



Sin embargo, la diferenciabilidad si implica la continuidad.

Teorema: Si una función es diferenciable en x_1 , entonces f es continua en x_1 .

5. Derivada de una Función

$$f'(x) = \lim_{Dx \rightarrow 0} \frac{f(x + Dx) - f(x)}{Dx}$$

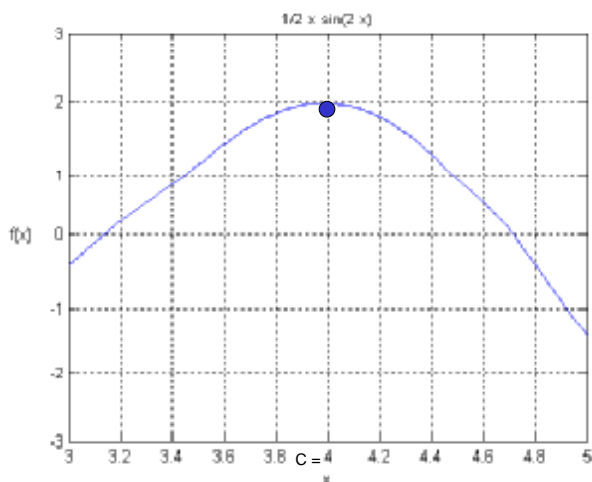
la pendiente de la recta tangente a $f(x)$ en el punto $(x, f(x))$.

6. Valores Máximos y Mínimos de una Función de una Variable

La derivada puede utilizarse para determinar los puntos donde la tangente es horizontal (derivada = 0).

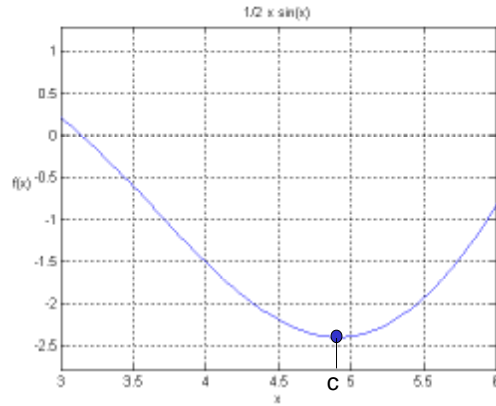
6.1. Extremos Relativos

Definición: La función f se dice que tiene un *valor máximo relativo* en “ c ”, si existe un intervalo abierto que contenga a “ c ” sobre el cual está definida la función f tal que $f(c) \geq f(x)$ para toda x en este intervalo.



Definición: La función f se dice que tiene un valor mínimo relativo en “ c ”, si existe un intervalo abierto que contenga a “ c ” sobre el cual f está definido tal que $f(c) \leq f(x)$ para toda x en este intervalo.

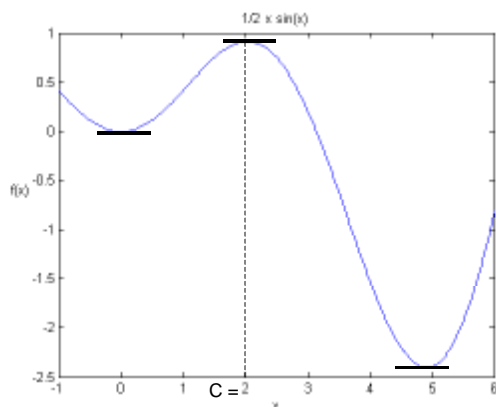
Ej.



¿Dónde Localizar los Posibles Valores Extremos?

Teorema: Si $f(x)$ existe para todos los valores de x en el intervalo abierto (a, b) y si f tiene un extremo relativo en “ c ”, donde $a < c < b$, si $f'(c)$ existe, $f'(c) = 0$.

Interpretación Geométrica



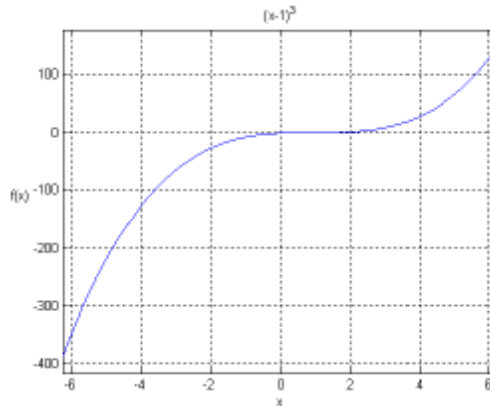
Extremos Relativos

$$f'(c) \text{ existe y } f'(c) = 0$$

Si f es una función diferenciable, los *únicos lugares posibles para puntos extremos* es donde $f'(x) = 0$.

Sin embargo, $f'(x)$ puede ser cero y no obstante en ese valor f no tiene un valor extremo (*Punto de Silla*).

Ej. $f(x) = (x - 1)^3$

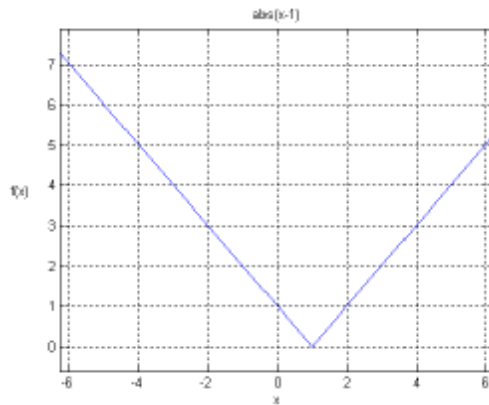


$$f'(x) = 3(x - 1)^2$$

$$f'(1) = 0$$

Más aún f puede tener un extremo relativo en un número y f' puede no existir allí.

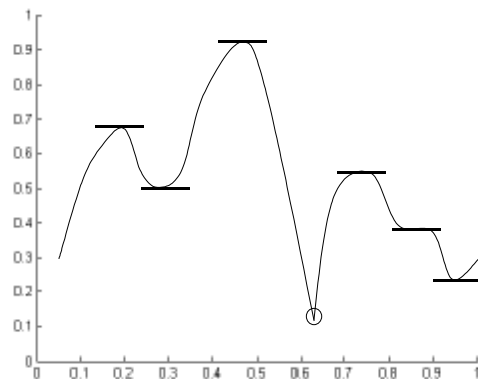
Ej. $f(x) = |x - 1|$



En Resumen,

Si una función está definida en un número “ c ” es una condición necesaria, pero no suficiente, para que f tenga un extremo relativo en “ c ” que $f'(c) = 0$ ó que $f'(c)$ no exista.

Definición: Si c es un número en el dominio de la función f y si $f'(c) = 0$ ó $f'(c)$ no existe, entonces “ c ” se llama *número crítico de f*.



6.2. Extremos Absolutos

Frecuentemente estamos en una función definida en un intervalo dado, y deseamos encontrar el valor mayor o menor de la función en el intervalo. Estos intervalos pueden ser cerrados, abiertos o cerrados a un extremo y abierto en otro.

El valor máximo absoluto es el mayor valor dentro del intervalo, y el valor mínimo absoluto es el mínimo valor de la función dentro del intervalo.

6.2.1. Extremos Absolutos en un Intervalo

Definición: La función f se dice que tiene un *valor máximo absoluto en un intervalo*, si existe algún número " c " en el intervalo tal que $f(c) \geq f(x)$ para toda x en el intervalo. En tal caso $f(c)$ es el valor máximo absoluto de f en el intervalo.

Definición: La función f se dice que tiene un *valor mínimo absoluto en un intervalo* si existe algún número " c " en el intervalo tal que $f(c) \leq f(x)$ para toda x en el intervalo. En tal caso $f(c)$ es el valor mínimo absoluto de f en el intervalo.

Valor extremo absoluto es un mínimo o máximo absoluto de función en el intervalo.

También se puede hablar de *extremo absoluto de una función* cuando no se especifica ningún intervalo.

Definición: Se dice que $f(c)$ es el *extremo máximo absoluto de la función f* si c está en el dominio de f y si $f(c) \geq f(x)$ para toda x en el dominio de la función.

Definición: $f(c)$ se dice que es un *mínimo global de la función f* si $c \in$ al dominio de f y si $f(c) \leq f(x) \forall x \in$ Dominio de f .

Teorema del Valor Extremo

Si una función f es continua en el intervalo cerrado $[a, b]$, entonces f tiene un valor máximo absoluto y un valor mínimo absoluto en $[a, b]$.

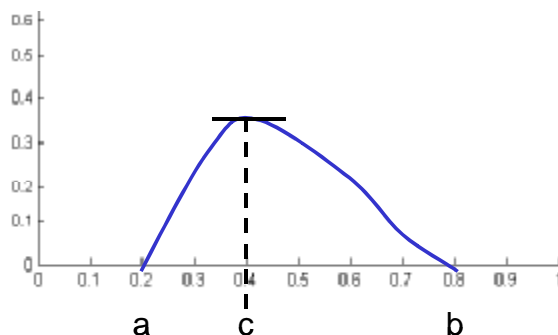
Un *extremo absoluto* de una función en un intervalo cerrado debe ser un *extremo relativo* o ser un *valor de la función en un extremo del intervalo*.

6.2.2. Procedimientos para determinación de extremos absolutos en intervalo cerrado:

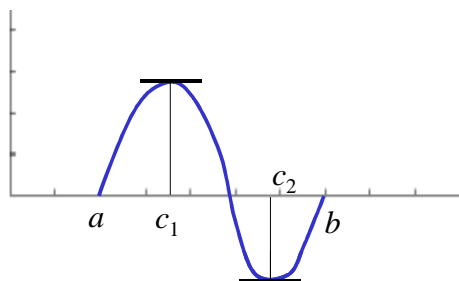
1. Identificar valores de la función en los números críticos de f en $[a, b]$
2. Encontrar $f(a)$ y $f(b)$
3. El *mayor* de estos es el *máximo absoluto* y el *menor* es el *mínimo absoluto*

Teorema de Rolle (Michel Rolle 1652-1719)

Sea una función f continua en un intervalo cerrado $[a, b]$, diferenciable en el intervalo (a, b) abierto y sean $f(a)=0$ y $f(b)=0$, existe al menos un número “ c ” entre a y b donde $f'(c)=0$.



Debe notarse que puede haber más de un número en el intervalo abierto (a, b) para el cual la derivada es cero.



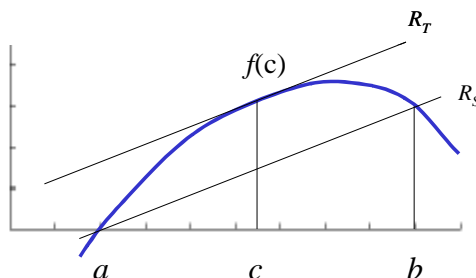
Teorema del Valor Medio (uno de los teoremas más importantes del cálculo diferencial e integral).

Sea f una función continua tal que:

- i. es continua en el intervalo cerrado $[a, b]$
- ii. es diferenciable en el intervalo abierto (a, b) .

entonces existe un número “ c ” en el intervalo abierto (a, b) tal que:

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$



La tangente R_T es paralela a la secante R_S

7. Funciones Crecientes y Decrecientes y Criterio de la Primera Derivada

Definición: Una función definida en un intervalo se dice que es *creciente* en ese intervalo si y solo si:

$$f(x_1) < f(x_2) \text{ siempre que } x_1 < x_2 \text{ donde } x_1 \text{ y } x_2 \text{ son números del intervalo}$$

Definición: Una función definida en un intervalo se dice que es *decreciente* en ese intervalo si y solo si:

$$f(x_1) > f(x_2) \text{ siempre que } x_1 < x_2 \text{ donde } x_1 \text{ y } x_2 \text{ son números del intervalo}$$

Si una función es creciente o decreciente en un intervalo, entonces se dice que *f es monótona*.

Teorema: Si una función continua en el intervalo cerrado $[a, b]$ y diferenciable en el intervalo abierto (a, b) .

- i. Si $f'(x) > 0$ para toda $x \Rightarrow$ es creciente en el intervalo
- ii. Si $f'(x) < 0$ para toda $x \Rightarrow$ es decreciente en el intervalo

7.1. Criterio de la Primera derivada para Extremos Relativos

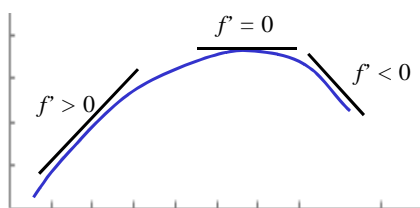
Si una función continua en el intervalo abierto (a, b) que contiene un número “ c ” y f es diferenciable, excepto, posiblemente en “ c ” ($f(c)$ puede no existir).

Si c es un extremo entonces:

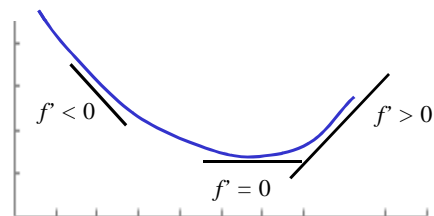
$$f(x_1) > f(c) \text{ donde } x_1 < c$$
$$f(x_2) < f(c) \text{ donde } c < x_2$$

en este caso c es un *máximo relativo*

Ej.



Máximo Relativo



Mínimo Relativo

7.1.1. Procedimiento para Determinar Extremos Relativos

1. Encontrar números críticos ($f'(x)=0$ ó $f'(x)$ no \exists)
2. Aplicar criterio de la primera derivada

7.2. Criterio de la Segunda Derivada

Sea “ c ” un número crítico de una función en la cual $f'(c)=0$ y f existe para todos los valores de x en algún intervalo abierto que contenga a “ c ”. Entonces si $f''(c)$ existe y,

- $i.$ Si $f''(c) < 0$, f tiene un valor máximo relativo en “ c ”.
- $ii.$ Si $f''(c) > 0$, f tiene un valor mínimo relativo en “ c ”.

Nótese que si $f''(c)=0$ nada puede concluirse.

Teorema: Sea una función continua en el intervalo I que contiene al número c . Si $f(c)$ es un extremo relativo de f en I y es el único, entonces $f(c)$ es un extremo absoluto de f en I . Además,

- $i.$ Si $f(c)$ es un máximo relativo \Rightarrow es un máximo absoluto
- $ii.$ Si $f(c)$ es un mínimo relativo \Rightarrow es un mínimo absoluto.

8. Formula de Taylor (Brook Taylor 1685 – 1731)

Ciertas funciones pueden ser aproximadas por polinomios y el polinomio puede ser usado cuando la diferencia es pequeña.

Teorema: Sea f una función tal que f y sus n primeras derivadas son continuas en el intervalo cerrado $[a, b]$. Además, $f^{(n+1)}(x)$ existe para toda x en el intervalo abierto (a, b) . Entonces hay un número e en el intervalo abierto (a, b) tal que,

$$f(b) = f(a) + \frac{f'(a)}{1!}(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(b-a)^n + \frac{f^{(n+1)}(e)}{(n+1)!}(b-a)^{n+1}$$

si $n=0 \Rightarrow f(b) = f(a) + f'(e)(b-a)$ [Teorema del Valor Medio]

8.1. Polinomio de Taylor

$$P_n(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2!} + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Residuo

$$R_n(x) = \frac{f^{(n+1)}(e)}{(n+1)!}(x-a)^{n+1} \text{ donde } e \text{ está entre } x \text{ y } a$$

Es buena aproximación cuando $x \cong a$ y cuando n es grande (relativamente grande).

9. Funciones de Varias Variables (Campos Escalares) $R^n \rightarrow R^1$

9.1. Continuidad de Campos Escalares

Definición: Sea la función $f = u \rightarrow R$ y sea $\bar{a} \in u$. Decimos que f es continua en \bar{a} si dado cualquier número positivo e , podemos hallar otro número positivo d tal que,

$$\text{Si } \|\bar{x} - \bar{a}\| < d, \bar{x} \in u, \text{ entonces } |f(\bar{x}) - f(\bar{a})| < e$$

ó

i. Si $f(\bar{a})$ está definida

ii. Si $\lim_{\bar{x} \rightarrow \bar{a}} f(\bar{x})$ existe \longrightarrow si esta falla entonces existe una discontinuidad esencial.

iii. Si $\lim_{\bar{x} \rightarrow \bar{a}} f(\bar{x}) = f(\bar{a})$ \longrightarrow si esta falla entonces existe una discontinuidad evitable.

Si f es continua en cada punto de u , decimos que f es continua en u .

9.2. Operaciones sobre Funciones Continuas

Si f y g son continuas en a entonces:

$$f + g, f - g \text{ y } f \times g \text{ son continuas}$$

$$f \div g \text{ es continua, si } g(\bar{a}) \neq 0$$

9.3. Gradiente

Recordando:

Derivada direccional: Razón de cambio en una dirección dada por un vector unitario \mathbf{u} .

Derivadas Parciales: Son derivadas direccionales pero especiales las direcciones son las de los ejes coordenados.

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \mathbf{L}$$

Definición: Si $f : u \rightarrow R, u \subseteq R^n$, la derivada de f en un punto $\bar{x}_0 \in u$ es el vector cuyos componentes son las derivadas parciales de f en \bar{x}_0 .

A esto se le llama *GRADIENTE*

$$\nabla f(\bar{x}_0) = \left(\frac{\partial f(\bar{x}_0)}{\partial x_1}, \frac{\partial f(\bar{x}_0)}{\partial x_2}, \mathbf{L}, \frac{\partial f(\bar{x}_0)}{\partial x_n} \right)$$

Definición: Sea $f : u \rightarrow R, u \subseteq R^n$ y $\bar{x} \in u$. Sea \mathbf{h} un vector que $\|\mathbf{h}\|$ sea adecuadamente pequeña de modo que $\bar{x} + \mathbf{h} \in u$, decimos que f es diferenciable en \bar{x} si existe el ∇f en \bar{x} y una función real $d(\mathbf{h})$ definida para esos valores de \mathbf{h} tales que:

$$f(\bar{x} + \mathbf{h}) - f(\bar{x}) = \nabla f(\bar{x}) \cdot \mathbf{h} + \|\mathbf{h}\| d(\mathbf{h})$$
$$\lim_{\mathbf{h} \rightarrow 0} \frac{d(\mathbf{h})}{\|\mathbf{h}\|} = 0$$

Teorema: Si $f : u \rightarrow R, u \subseteq R^n$, es diferenciable en \bar{x} entonces es continua en \bar{x} .

El recíproco es falso: Una función puede ser continua sin ser diferenciable.

Teorema: (Condición de *suficiencia* de diferenciability).

Si $f : u \rightarrow R, u \subseteq R^n$ posee derivadas parciales continuas en $\bar{x}_0 \in u$ entonces f es diferenciable en \bar{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 $u \subseteq R^n$ y escribimos $f \in C^k(u)$, si todas sus derivadas parciales de orden k existen y son continuas en u .

Si $f \in C^k(u)$ esto implica la continuidad de las derivadas de orden menor en u .

9.4. Formula de Taylor en Varias Variables

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{j=1}^h \frac{\partial f(x_0)}{\partial x_j} (x_j - x_{0j}) + \frac{1}{2!} \sum_{i=1}^h \sum_{j=1}^h (x_i - x_{0i})(x_j - x_{0j}) \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} + \mathbf{L}$$

En una notación más convencional y compacta donde:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \mathbf{L}, \frac{\partial f}{\partial x_n} \right) \text{ es el gradiente y,}$$

$$H \equiv \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \mathbf{L} & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \mathbf{L} & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \mathbf{L} & \frac{\partial^2 f}{\partial x_n^2} \end{vmatrix} \text{ es la matriz Hessiana}$$

la formula de Taylor de varias variables es,

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f^T(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2!} (\mathbf{x} - \mathbf{x}_0)^T \bar{H} (\mathbf{x} - \mathbf{x}_0) + \mathbf{K}$$

Los términos de mayor orden no se escriben por su complejidad y porque no serán necesarios en nuestro trabajo en “no linear programming”.

9.5. Extensión de los Criterios de Existencia de Máximo y Mínimos

- Los puntos críticos son aquellos donde $\nabla f = 0$ o no existe.
- Alguna medida de “positividad” del Hessiano nos dirá si es un máximo o un mínimo.

¿Por qué no resolver el problema?

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial x_1} = 0 \\ \frac{\partial f}{\partial x_2} = 0 \\ \mathbf{M} \\ \frac{\partial f}{\partial x_n} = 0 \end{array} \right. \quad \text{Porque es más complejo que los métodos que vamos a estudiar}$$

Teorema de Weierstrass (Extensión del teorema de Valor Extremo):

Una función continua f , definida en un conjunto compacto S cerrado y acotado (definido y no se va a infinito) tiene un mínimo y un máximo en S .

10. Formas Cuadráticas

A pesar que en ingeniería, economía y estadística la mayoría de los problemas son representados (frecuentemente de forma no apropiada) por modelos lineales y resuelto por métodos lineales, los modelos cuadráticos y técnicas cuadráticas están en segundo lugar.

Definición: Una forma cuadrática es cualquier campo escalar ($R^n \rightarrow R$), definido para todo x en R^n que sigue la siguiente forma:

$$q(\mathbf{x}) = \sum_{i=1}^h \sum_{j=1}^h a_{ij} x_i x_j \quad \text{donde } a_{ij} \in R \text{ puede ser cero (*)}$$

Una forma cuadrática no incluye ningún término lineal.

Ej. $q(x_1, x_2) = 7x_1^2 + 4x_1x_2 + 8x_2^2 \neq \text{función cuadrática}$

Notación Matricial

Cualquier forma cuadrática puede ser expresada como:

$$q(x) = x^t A x \quad \text{donde } a_{ij} \text{ son elementos de la matriz } A \text{ y son los coeficientes en (*)}$$

Es claro que:

$$a_{ij} x_i x_j + a_{ji} x_j x_i = (a_{ij} + a_{ji}) x_i x_j \quad \text{para todo } i \neq j$$

Por lo tanto una forma cuadrática puede ser representada equivalentemente por muchas matrices A o conjuntos de coeficientes a_{ij} .

Sin embargo, para una forma cuadrática $q(\mathbf{x})$ dada existe solo una matriz simétrica (cuadrada tal que $D = D^t$) que satisface $q(\mathbf{x}) = \mathbf{x}^t D \mathbf{x}$ cuyos elementos están definidos por:

$$d_{ij} = d_{ji} = \frac{(a_{ij} + a_{ji})}{2} \quad \text{para todo } i \text{ y } j$$

Ej. $q(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2$

$$q(x_1, x_2) = x^t D x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Ej. Si $A = \begin{vmatrix} 2 & -2 \\ 8 & -1 \end{vmatrix}$

$$q(x_1, x_2) = 2x_1^2 - 2x_1x_2 + 8x_2x_1 - x_2^2 = 2x_1^2 + 6x_1x_2 - x_2^2$$

$$q(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

10.1. Propiedades de las formas cuadráticas

Definición: La forma cuadrática $q(\mathbf{x}) \equiv \mathbf{x}' D \mathbf{x}$ es definida positiva si $q(\mathbf{x}) > 0$ para todo $\mathbf{x} \neq 0$ en E^n .

Definición: La forma cuadrática $q(\mathbf{x}) \equiv \mathbf{x}' D \mathbf{x}$ es semidefinida positiva si $q(\mathbf{x}) \geq 0$ para todo $X \in E^n$, pero $q(\mathbf{x})$ no es definida positiva.

Definición: La forma cuadrática $q(\mathbf{x}) \equiv \mathbf{x}' D \mathbf{x}$ es definida negativa si $q(\mathbf{x}) < 0$ para todo $\mathbf{x} \neq 0$ en E^n .

Definición: La forma cuadrática $q(\mathbf{x}) \equiv \mathbf{x}' D \mathbf{x}$ es semidefinida negativa si $q(\mathbf{x}) \leq 0$ para todo $X \in E^n$, pero $q(\mathbf{x})$ no es definida negativa.

La matriz D (única y simétrica) de una forma cuadrática definida positiva es definida positiva.

Si no satisface ninguna de las cuatro definiciones anteriores se dice que la forma cuadrática es indefinida. Esto es si $q(\mathbf{x}_1) > 0$ y $q(\mathbf{x}_2) < 0 \Rightarrow q(\mathbf{x})$ es indefinida, donde \mathbf{x}_1 y $\mathbf{x}_2 \in R^n$.

Ej: $q(x_1, x_2) = x_1^2 + 10x_2^2$ es definida positiva

$q(x_1, x_2) = -x_1^2 + (-10)x_2^2$ es definida negativa

$q(x_1, x_2, x_3) = 2(x_1 - 2x_2) + x_3^2$ es semidefinida positiva porque $q(2, 1, 0) = 0$

Propiedad: Si $q(\mathbf{x})$ es definida positiva (o semi definida positiva) entonces $-q(\mathbf{x})$ es definida negativa (o semi definida negativa) y viceversa.

Lema: (IMPORTANTE!!!) útil en *Quadratic Programming*.

Sea D una matriz simétrica de $n \times n$ definida positiva (negativa). Entonces:

- a) D^{-1} existe
- b) D^{-1} es definida positiva (negativa)
- c) $AD A^t$ es semidefinida positiva (negativa) para cualquier matriz $A m \times n$.

10.2. Métodos de clasificación de formas cuadráticas

10.2.1. Método de los menores principales

Sea $q(\mathbf{x}) \equiv \mathbf{x}^t D \mathbf{x}$ una forma cuadrática, con D matriz simétrica. Sean $|D_1|, |D_2|, \dots, |D_n|$ los n menores principales de la matriz D , donde

$$\begin{aligned}
 |D_1| &= |d_{11}| \\
 |D_2| &= \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} \\
 |D_n| &= |D|
 \end{aligned}$$

es decir, los n determinantes de las n submatrices de D formadas por las i primeras filas y las i primeras columnas, para $i = 1 \dots n$.

Entonces:

- a) $q(\mathbf{x})$ es definida positiva si y sólo si todos los menores principales de D son positivos ($|D_1| > 0, |D_2| > 0, \dots, |D_n| > 0$).
- b) $q(\mathbf{x})$ es definida negativa si y sólo los menores principales de D alternan de signo empezando por negativo ($|D_1| < 0, |D_2| > 0, \dots, |D_n|$ con el signo de $(-1)^n$).
- c) Si los $n-1$ primeros menores principales son todos positivos y el último $|D_n| = 0$, entonces $q(\mathbf{x})$ es semidefinida positiva.
- d) Si los $n-1$ primeros menores principales alternan de signo empezando por negativo y el último $|D_n| = 0$, entonces $q(\mathbf{x})$ es semidefinida negativa.
- e) Si $|D_n| = |D| \neq 0$ y los criterios a) y b) no se verifican, entonces $q(\mathbf{x})$ es indefinida.
- f) Si $|D_i| \neq 0$ para $i = 1, 2, \dots, n-1$, $|D_n| = |D| = 0$ y los criterios c) y d) no se verifican, entonces $q(\mathbf{x})$ es indefinida.

10.2.2. Método de los autovalores:

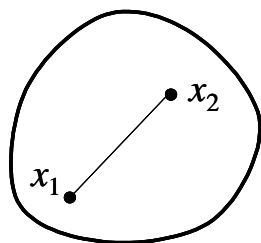
Sea $q(\mathbf{x}) \equiv \mathbf{x}^t D \mathbf{x}$ una forma cuadrática, con D matriz simétrica. Sean $\lambda_1, \lambda_2, \dots, \lambda_n$ los n autovalores de la matriz D . Entonces:

- a) $q(\mathbf{x})$ es definida positiva si y sólo si $\lambda_i > 0 \forall i$.
- b) $q(\mathbf{x})$ es definida negativa si y sólo si $\lambda_i < 0 \forall i$.
- c) $q(\mathbf{x})$ es semidefinida positiva si y sólo $\lambda_i \geq 0 \forall i$, siendo al menos un $\lambda_j = 0$.
- d) $q(\mathbf{x})$ es semidefinida negativa si y sólo $\lambda_i \leq 0 \forall i$, siendo al menos un $\lambda_j = 0$.
- e) $q(\mathbf{x})$ es indefinida si y sólo si algún $\lambda_i > 0$ y algún $\lambda_j < 0$.

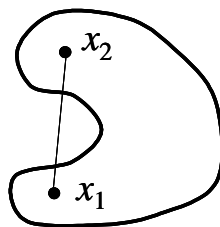
11. Funciones Convexas

Estamos particularmente interesados en la optimización de este tipo de funciones sobre los llamados *conjuntos convexos*.

Definición: Un conjunto X en E^n (R^n) es convexo si y sólo si para dos puntos cualquiera \mathbf{x}_1 y \mathbf{x}_2 en X y cualquier valor escalar $0 \leq I \leq 1$, el punto $\mathbf{x} = I\mathbf{x}_1 + (1 - I)\mathbf{x}_2$ también está dentro de X .



Convexo

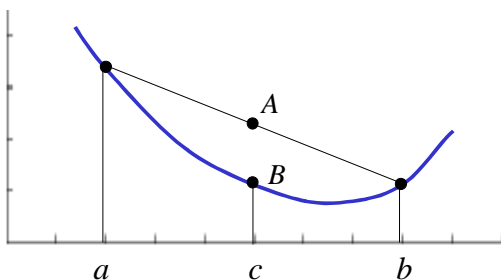


No Convexo

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

Definición: Una función escalar $f(\mathbf{x})$ es una función convexa definida sobre un conjunto convexo X en E^n si para dos puntos cualquiera \mathbf{x}_1 y \mathbf{x}_2 en X .

$$f(I\mathbf{x}_1 + (1 - I)\mathbf{x}_2) \leq If(\mathbf{x}_1) + (1 - I)f(\mathbf{x}_2) \text{ donde } 0 \leq I \leq 1$$



$$\overline{AC} = If(a) + (1 - I)f(b)$$

$$\overline{BC} = f(Ia + (1 - I)b)$$

Las funciones convexas tienen una caracterización geométrica simple e informativa.

Teorema: Cualquier función lineal $f(\mathbf{x}) = \mathbf{c}'\mathbf{x}$ es tanto cóncava como convexa.

Teorema: Si $f(\mathbf{x})$ es convexa $\Rightarrow -f(\mathbf{x})$ es cóncava (y viceversa).

Teorema: La suma de 2 o más funciones convexas (cóncavas) es convexa (cóncava)

Teorema: Cualquier forma cuadrática semidefinida positiva $q(\mathbf{x}) = \mathbf{x}'D\mathbf{x}$ donde D es simétrica, es una función convexa en todo E^n . Si es definida positiva es estrictamente convexa.

Teorema: Cualquier forma cuadrática semidefinida negativa $q(\mathbf{x}) = \mathbf{x}'D\mathbf{x}$ donde D es simétrica, es una función cóncava en todo E^n . Si es definida negativa es estrictamente cóncava.

Dada una función cuadrática representada como $f(\mathbf{x}) \equiv q(\mathbf{x}) + \mathbf{c}'\mathbf{x} + c_0$
es convexa o cóncava si $q(\mathbf{x})$ es convexa o cóncava.

¿A que se parece?

Al desarrollo de Taylor

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f'(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)' H(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

Teorema: Si la función $f(\mathbf{x})$ está definida y es convexa sobre un conjunto convexo X en E^n , luego cualquier mínimo local (con restricción) de $f(\mathbf{x})$ en X es un mínimo global en X .

Idem para máximo-concava

Tiene una implicación importante: el primer extremo identificado es la solución deseada.

Teorema: Si una función $f(\mathbf{x})$ es convexa (cóncava) sobre un conjunto X compacto y convexo (cerrado y limitado) en E^n entonces al menos un máximo (mínimo) global se encuentra sobre el borde de X .

11.1. Caracterización alternativa de las funciones convexas.

En general es necesario determinar si una función es convexa. Hasta ahora se han derivado resultados para funciones lineales y cuadráticas.

Para cierta cantidad de funciones (no tan complejas) de algunas variables es posible determinar su convexidad mediante una prueba de la segunda derivada.

11.1.1. Criterios de la primera y segunda derivada

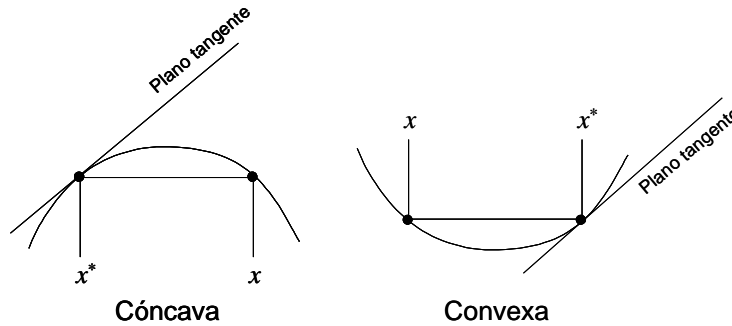
Teorema: Supongamos que $f(\mathbf{x})$ tiene primeras derivadas parciales continuas. Luego $f(\mathbf{x})$ es cóncava sobre alguna región R en E^n si y sólo si

$$f(\mathbf{x}) \leq f(\mathbf{x}^*) + \nabla f^t(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*)$$

similarmente, $f(\mathbf{x})$ es convexa sobre alguna región R en E^n si y sólo si

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f^t(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*)$$

Aquí se requiere que un hiperplano que es tangente en algún punto x^* a una hipersuperficie cóncava debe estar enteramente sobre ella.



Teorema: Sea $f(\mathbf{x})$ una función $\in C^2$ (segundas derivadas parciales existen y son continuas). Entonces $f(\mathbf{x})$ es cóncava (convexa) sobre una región R en E^n si y sólo si su Hessiano es definido o semi definido negativo (positivo) para toda \mathbf{x} de la región R .

Teorema de Schwartz

Si $f(\mathbf{x}, y)$ es tal que $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ y $\frac{\partial^2 f}{\partial x \partial y}$ son continuas en un entorno de un punto (x_0, y_0) ,

entonces $\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$ existe y se cumple que $\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$

(Debido al teorema de Schwartz)

Como la matriz Hessiano es simétrica la definición definida y semidefinida positiva para formas cuadráticas es aplicable directamente.

Una función puede ser convexa o cóncava y su Hessiano puede “desaparecer” en algunos puntos (matriz de ceros).