CAPÍTULO 2

PROGRAMACIÓN NO LINEAL

Capítulo 2: Programación No Lineal

$$\text{mín} \ \ (\text{ó} \ \ \text{máx}) f(x)$$
$$\text{s.a. } x \in S \subseteq R^n$$

No existe un método que permita resolver cualquier problema de programación no lineal. Los métodos existentes sólo resuelven algunos tipos de problemas. Veremos conceptos y herramientas que proveen el marco teórico para la resolución de problemas con ciertas condiciones.

Conceptos Básicos

1. **Conjuntos Convexos:**
   Dados los vectores $x^1, x^2, \ldots, x^k \in R^n$ se dice que el vector y es una combinación lineal convexa de $x^1, x^2, \ldots, x^k$ si:

$$y = \sum_{i=1}^{k} \alpha_i x^i, \text{ con } \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i = 1$$

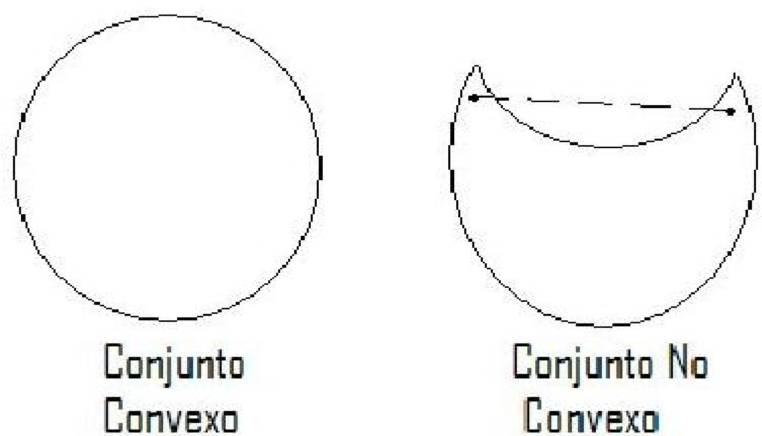
Conceptos Básicos

Definición:
Un conjunto $S \subseteq R^n$ se dice convexo si el segmento de recta que une 2 puntos cualesquiera de S está totalmente contenido en S. O sea, $\forall x^1; x^2 \in S$ se tiene que:

$$\lambda x^1 + (1-\lambda)x^2 \in S, \forall \lambda \in [0,1]$$

Esto corresponde a una combinación lineal convexa de x^1, x^2

Ej:

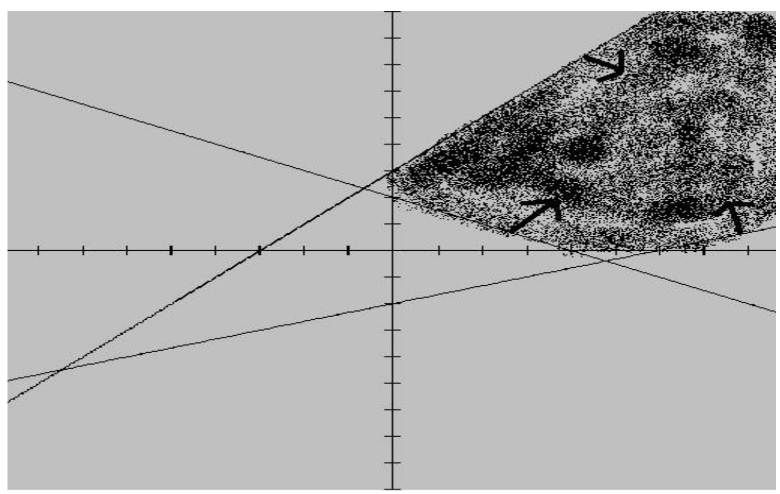


Conceptos Básicos

Teorema:
Sean $S_1, S_2, \ldots, S_p$ conjuntos convexos. Entonces $S = \bigcap_{i=1}^{n} S_i$ también es convexo.

Observación: Por convención el conjunto ϕ es convexo.

Ej:
El conjunto $S = \{(x_1, x_2)/x_1 - 3x_2 \leq 6, x_1 + 2x_2 \geq 4, 2x_1 - 2x_2 \geq -6\}$ es convexo (es la intersección de los 3 semiespacios definidos por las respectivas desigualdades lineales).

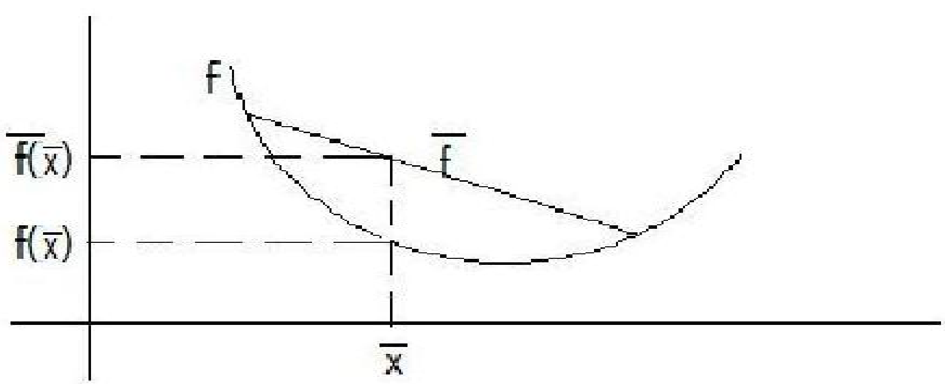


Conceptos Básicos

2. **Funciones Convexas:**
Definición:
Sea $f : S \subseteq R^n \longrightarrow R$ una función escalar, donde S es un conjunto convexo no vacío. f se dice convexa en S si:

$$f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda)f(x^2), \forall x^1, x^2 \in S, \forall \lambda \in [0,1]$$

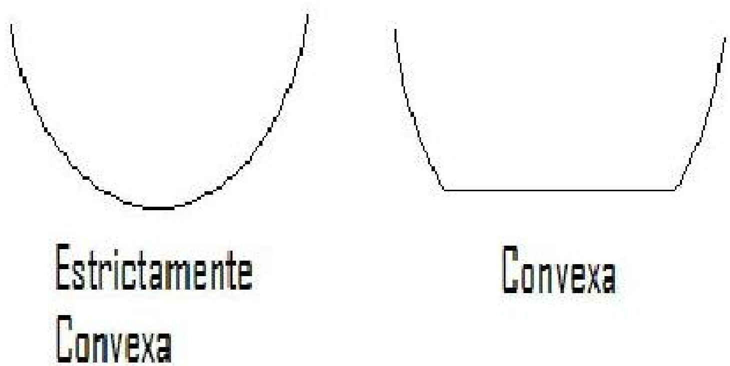
Interpretación Geométrica:
Una función es convexa cuando su interpolación lineal $\bar{f}$ entre 2 puntos no subestima el valor de la función.



Conceptos Básicos

Definición:
La función f es **estrictamente convexa** en S si la desigualdad anterior es estricta para todo $x^1 \neq x^2 \in S$ y $\forall \lambda \in (0,1)$.

Ej:



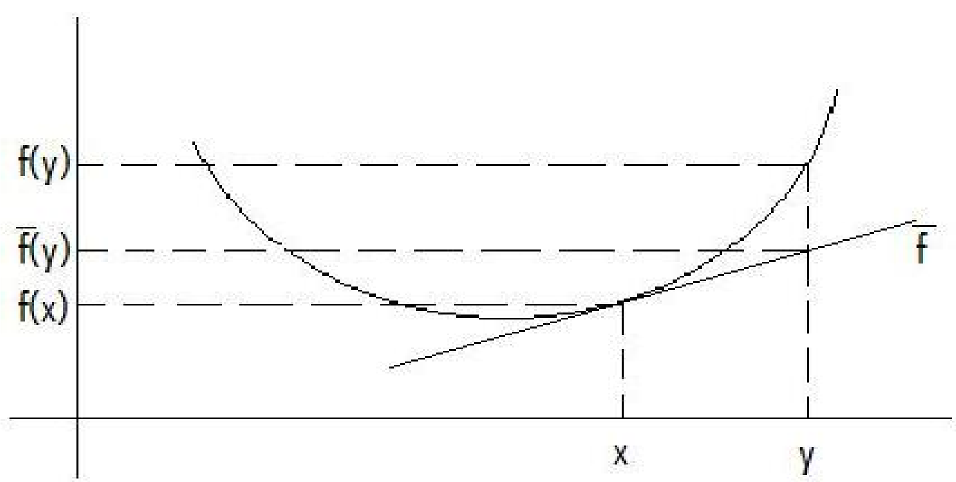
- Estrictamente Convexa:
  $f(x,y) = x^4 + y^2$
- Convexa:
  $f(x) = sx$

Conceptos Básicos

Teorema:
Sea la función $f \in C^1$. Entonces f es convexa en el conjunto convexo S sii:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \forall x, y \in S$$

Interpretación Geométrica:
Dado un cierto valor de x, la función $\overline{f}(y) = f(x) + \nabla f(x)^T (y - x)$ corresponde a la aproximación lineal de f basada en la derivada direccional de f en x.
El teorema establece que esta aproximación lineal nunca supera el valor de una función convexa.



Conceptos Básicos

Ej: $f(x) = (x-3)^2 + 2$

$$\overline{f}(y) = (x-3)^2 + 2 + 2(x-3)(y-x)$$

$$\begin{gathered}\overline{f}(y) = -x^2 + 2xy - 6y + 11 + y^2 - y^2 \\ \overline{f}(y) = -(x-y)^2 + (y-3)^2 + 2 \\ \Rightarrow \overline{f}(y) \leq (y-3)^2 + 2 = f(y), \forall x, y \\ \therefore f \text{ convexa por el teorema}\end{gathered}$$

Teorema:
Sea $f \in C^2$ y sea $S \subseteq R^n$ un conjunto convexo, abierto y no vacío.
Entonces f es convexa si y sólo si $Hf(x)$ es semidefinida positiva en todo S.

Recordemos como se calcula la matriz hessiana: Sea $f(x,y) \in C^2$, la matriz hessiana $Hf(x,y)$ viene dada por:

$$\left[\begin{array}{cc} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{array}\right]$$

Conceptos Básicos

Teorema:
Sea $f \in C^2 \Rightarrow \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}, \forall x, y$.

Corolario:
$f \in C^2 \Rightarrow Hf(x)$ es simétrica.

Repaso de Matrices Definidas Positivas

Definición:
Sea $A \in R^{n \times n}$ una matriz simétrica. Se dice que A es definida positiva si la forma cuadrática $x^T Ax$ es positiva para todo vector x diferente al vector nulo, es decir, si $x^T Ax > 0, \forall x \neq 0$.

Definición:
A es semidefinida positiva si $x^T Ax \geq 0, \forall x \neq 0$.

Definición:
A es definida negativa si $x^T Ax < 0, \forall x \neq 0$

Definición:
A es semidefinida negativa si $x^T Ax \leq 0, \forall x \neq 0$

Observación: Si $x^T Ax > 0$ para algún $x \in R^n$ y $x^T Ax < 0$ para otro $x \in R^n$, entonces se dice que A es indefinida.

Repaso de Matrices Definidas Positivas

Relación con los Valores Propios:
A es definida (semidefinida) positiva $\Leftrightarrow$ todos los Valores Propios de A son positivos (no negativos).

Relación con los Subdeterminantes de la Matriz:
Definición:
Una matriz A simétrica es definida positiva $\Leftrightarrow$ el determinante de cada una de sus submatrices principales es positivo.

Definición:
Una matriz A simétrica es definida negativa $\Leftrightarrow$ el determinante de la k-ésima submatriz principal de A tiene signo $(-1)^k, k = 1, \ldots, n$.

Observación: Esta caracterización no se puede extender a matrices semidefinidas positivas o negativas.

Ej:

a) $f(x,y) = x^2 - 2xy + y^2$
El Hessiano Hf(x) es:

$$\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$$

Veamos que tipo de matriz es:

$$(x\ y) \cdot Hf(x) \cdot (x\ y)^T = 2(x-y)^2 \geq 0, \forall (x\ y)$$

$\therefore Hf(x)$ es semidefinida positiva $\Rightarrow f$ es convexa.

b) $f(x,y) = (x^2 + y^2)^2$
El Hessiano Hf(x) es:

$$\begin{bmatrix} 12x^2 + 4y^2 & 8xy \\ 8xy & 12y^2 + 4x^2 \end{bmatrix}$$

Veamos que tipo de matriz es:

$$(x\ y) \cdot Hf(x) \cdot (x\ y)^T = 12x^4 + 24x^2y^2 + 12y^4 > 0, \forall (x\ y) \neq (0\ 0)$$

$\therefore Hf(x)$ es definida positiva $\Rightarrow f$ es convexa.

Conceptos Básicos

Teorema:
Sea S un conjunto convexo, abierto y no vacío, y sea $f \in C^2$. Si la matriz hessiana de f es definida positiva en todo punto de $S \Rightarrow f$ es estrictamente convexa.

Observación: La recíproca del Teorema no es necesariamente cierta. Si f es estrictamente convexa $\Rightarrow Hf(x)$ es semidefinida positiva (y no necesariamente definida positiva).

Ej:
$f(x) = x^6 \Rightarrow Hf(x) = 30x^4$ es definida positiva $\forall x \neq 0$ pero es semidefinida positiva en $x = 0$.

Propiedades de las Funciones Convexas

$S \subseteq R^n$ **convexo y no vacío** y $f : S \longrightarrow R$ una función escalar.

a) Sean $f_1, f_2, \ldots, f_k : S \longrightarrow R$ funciones convexas. Entonces la función $f(x) = \sum_{i=1}^{k} \alpha_i f_i(x)$ con $\alpha_i \geq 0$ para $i = 1, \ldots, k$ es también convexa. La función $f(x) = \max\{f_1, \ldots, f_k\}$ es convexa.

b) **Definición:** La función $g : S \longrightarrow R$ es cóncava si la función $-g(x)$ es convexa.
Observación: La función lineal $f(x) = \alpha^T x + \beta$ es convexa y cóncava a la vez.

c) Sea f una función convexa y $g : R \longrightarrow R$ una función convexa no decreciente $\Rightarrow h(x) = g[f(x)]$ es convexa.

d) El epígrafo de S, denotado epi(f) es un subconjunto de R^{n+1} definido por:

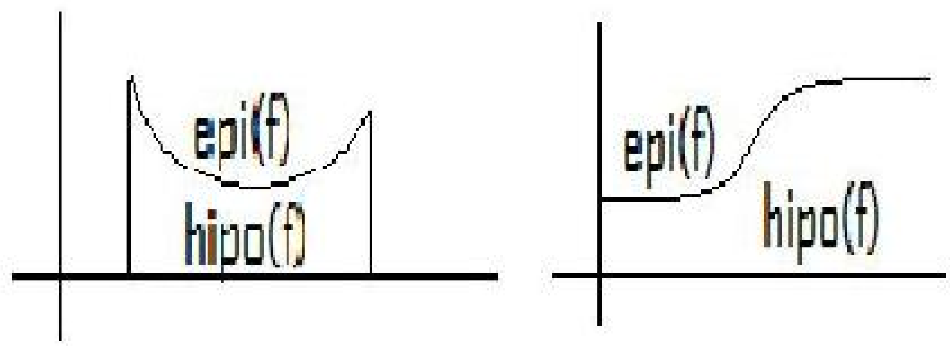
$$\mathrm{epi}(f) = \{(x, y)/x \in S, y \in R, y \geq f(x)\}$$

Análogamente, el hipógrafo de f (hipo(f)) es:

$$\mathrm{hipo}(f) = \{(x, y)/x \in S, y \in R, y \leq f(x)\}$$

Propiedades de las Funciones Convexas

Propiedad: f es convexa $\Leftrightarrow$ epi(f) es un conjunto convexo



$e)$ Si f es convexa, entonces f es continua en el interior de S.

Observación: Se pueden establecer propiedades análogas a las anteriores para las funciones cóncavas.

Conceptos Básicos

3. **Óptimos Locales y Globales:**

$$(\mathrm{P})\, \min f(x)$$

$$\text{s.a. } x \in S$$

Un $x \in S$ se denomina solución factible del problema.

Mínimo Global:
Sea $S \subseteq R^n$ y $f : S \longrightarrow R^n$. Un punto $\overline{x} \in R$ se dice **mínimo global** o **mínimo** de (P) si:

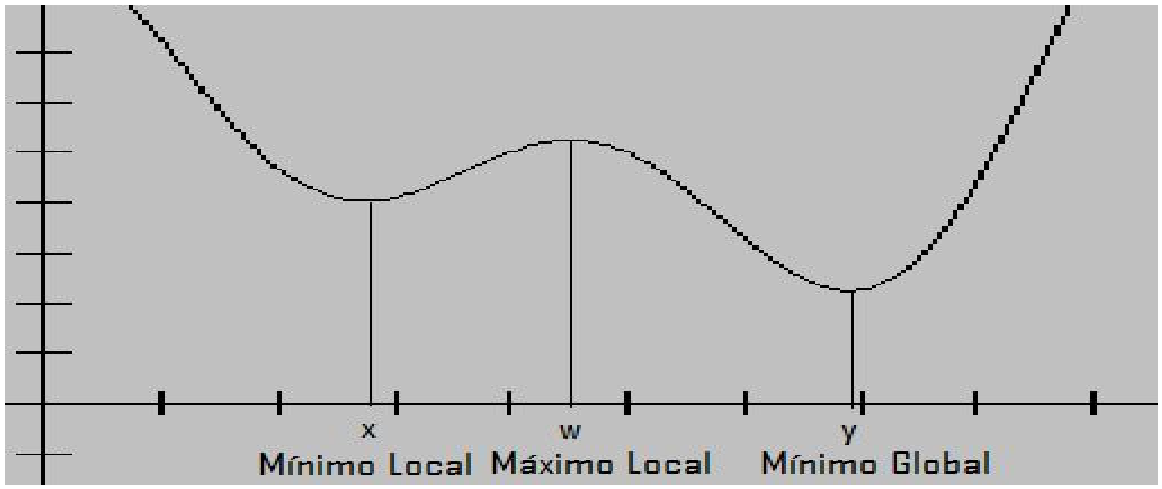
$$\overline{x} \in S \text{ y } f(\overline{x}) \leq f(x), \forall x \in S$$

Mínimo Local:
Un punto $\overline{x} \in R^n$ se dice mínimo local de (P) si:

$$\overline{x} \in S \text{ y } \exists \delta / f(\overline{x}) \leq f(x), \forall x / \|x - \overline{x}\| \leq \delta$$

Conceptos Básicos

Definición:
El valor que toma la función f en el mínimo global se denomina valor óptimo.



A continuación el estudio se centrará en encontrar puntos óptimos. Para esto se separarán dos tipos de problemas:

1. Optimización Irrestricta

2. Optimización con restricciones

Optimización Irrestricta

$$(\mathrm{P})\ \min f(x)$$
$$\text{s.a. } x \in R^n$$

Nota: $\max f = \min(-f)$

Caracterización de Puntos Óptimos:
Definición:
Si $f : R^n \longrightarrow R$ es diferenciable sobre R^n, un punto $\overline{x}$ se dice estacionario para f si $\nabla f(\overline{x}) = 0$.

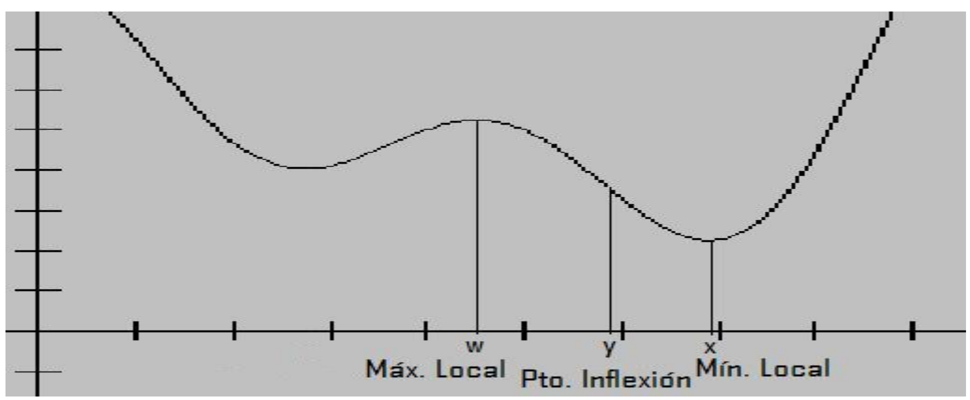
Teorema: (Condición Necesaria)
Sea $f : R^n \longrightarrow R$ una función diferenciable. Si $\overline{x}$ es mínimo local de f, entonces $\nabla f(\overline{x}) = 0$. (todo punto mínimo local es estacionario)

Optimización Irrestricta

Demostración: $(-q \Rightarrow -p)$
Sea $\overline{x}$ que no cumple $\nabla f(\overline{x}) = 0$ y sea $d = -\nabla f(\overline{x})$. Entonces se tiene que $d^T \nabla f(\overline{x}) < 0$. Luego tenemos que f decrece localmente alrededor de $\overline{x}$ en la dirección de d, contradiciendo el hecho de que $\overline{x}$ es mínimo local.

Observación: Condición Necesaria, pero no suficiente.

Ej: $f(x) = 2x^5 + 3$
$f'(x) = 10x^4 \longrightarrow x = 0$ es estacionario pero no es mínimo local sino punto de inflexión.



Optimización Irrestricta

Para dar una condición suficiente, necesitamos información de segundo orden de la función, lo cual se obtiene a través de $Hf(x)$.

Teorema: (Condición Suficiente)
Sea $f : R^n \longrightarrow R$ una función C^2 y sea $\overline{x}$ tal que $\nabla f(\overline{x}) = 0$. Entonces si $Hf(\overline{x})$ es definida positiva, $\overline{x}$ es mínimo local de f.

Demostración:
Hagamos el desarrollo de Taylor en torno a $\overline{x}$. Sea h pequeño y mayor que 0.

$$f(\overline{x} + h) = f(\overline{x}) + h^T \nabla f(\overline{x}) + \frac{1}{2} h^T Hf(\overline{x}) h + R(h)$$

Donde $R(h)$ es tal que:

$$\lim_{h \longrightarrow 0} \frac{|R(h)|}{\|h^2\|} = 0$$

Luego vemos que:

$$h^T \nabla f(\overline{x}) = 0 \text{ ya que } \nabla f(\overline{x}) = 0$$

Optimización Irrestricta

$$\frac{1}{2}h^T Hf(\overline{x})h > 0 \text{ ya que } Hf(\overline{x}) \text{ es definida positiva}$$
$$\Rightarrow f(\overline{x}+h) > f(\overline{x})$$
$$\therefore \overline{x} \text{ mínimo local.}$$

Observación: Si el Hessiano es indefinido o semidefinido positivo, no es posible concluir nada, podría tratarse de un *punto de silla*.

Ej: $f(x,y) = x^3 + y^2$
$(0,0)$ punto estacionario. $Hf(0,0)$ es:

$$\begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

Vemos que $Hf(0,0)$ es semidefinida positiva y el punto $(0,0)$ es un punto de inflexión.

Optimización Irrestricta

Observación: Una función f puede admitir mínimos que no satisfagan las condiciones enunciadas anteriormente.

Ej: $f(x) = x^{\frac{2}{3}}$
Mínimo: $x = 0$ pero $\nabla f(x)$ no está definido en el 0.

Caso Hessiano Indefinido

Es posible mostrar que si $Hf(x)$ es indefinida pero diagonalizable (así será si la función es 2 veces continuamente diferenciable), entonces podemos decir más sobre f:

Sea $Q \in R^{n\times n}$ y $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ tal que $Q^T Hf(x)Q = D$. El signo de los valores propios $\lambda_1, \ldots, \lambda_n$ da información sobre f alrededor de x: (recordemos que si $\lambda_i > 0 \Rightarrow Hf(x)$ definida positiva)

1. Si $\lambda_{i0} > 0$, sea v_{i0} el vector propio respectivo. Entonces v_{i0} es una dirección de crecimiento de la función.

2. Si $\lambda_{i1} < 0$, sea v_{i1} el vector propio respectivo. Entonces v_{i1} es una dirección de decrecimiento de la función.

3. Si pasan 1 y 2, nos encontramos en un punto silla.

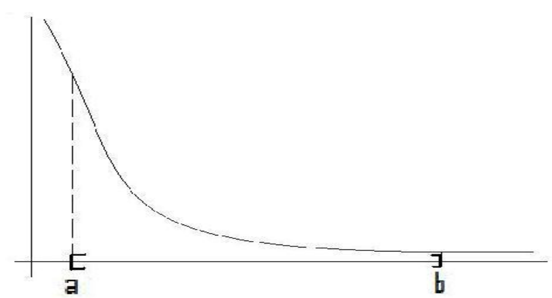
El Caso Convexo

Si se agrega la hipótesis de convexidad de la función el análisis se simplifica.

Sea $f : S \subseteq R^n \longrightarrow R$ una función escalar convexa sobre S, convexo y no vacío. Así, las propiedades vistas anteriormente también sirven para optimización con restricciones. Luego:

1. Si $\overline{x}$ es mínimo local $\Rightarrow$ es mínimo global.

2. Sea f convexa y diferenciable en $\overline{x}$ punto interior de S. Entonces $\overline{x}$ es óptimo global de f en $S \Leftrightarrow \nabla f(\overline{x}) = 0$.
   Ej:



El Caso Convexo

El mínimo en el intervalo cerrado $[a, b]$ es el punto b que no es interior y donde ∇f es no nulo.

3. Si f es estrictamente convexa, admite a lo más un punto mínimo global.

4. Si el problema de optimización irrestricto admite soluciones, entonces el conjunto de soluciones es convexo.

5. Si la función f es diferenciable en $\overline{x} \in S$, entonces $\overline{x}$ es mínimo global de $f \Leftrightarrow \nabla f(\overline{x})^T(x - \overline{x}) \geq 0, \forall x \in S$

Ejercicios: Probar usando propiedades de funciones y conjuntos convexos.

Métodos de Descenso para Funciones Diferenciables

Dado lo difícil que es buscar puntos estacionarios en una función de forma analítica, se desarrollarán métodos iterativos que determinarán estos puntos por medio de aproximaciones sucesivas.

Se busca un punto estacionario de una función escalar f. Se generan iterativamente una sucesión de puntos $\{x^k\}, k = 0, 1, 2, \ldots$ tal que la sucesión correspondiente $\{f(x^k)\}, k = 0, 1, 2, \ldots$ sea monótona decreciente. Una sucesión que verifica esta propiedad se dice que es una *sucesión de descenso* para f.

Definición: Sea $f : R^n \longrightarrow R$ una función diferenciable. Se llama **dirección de descenso** para f en $\overline{x}$ a todo vector $d \in R^n$ tal que $d^T \nabla f(\overline{x}) < 0$. Esto quiere decir que el vector d forma un ángulo agudo con el vector $-\nabla f(\overline{x})$, disminuyendo el valor de la función objetivo.

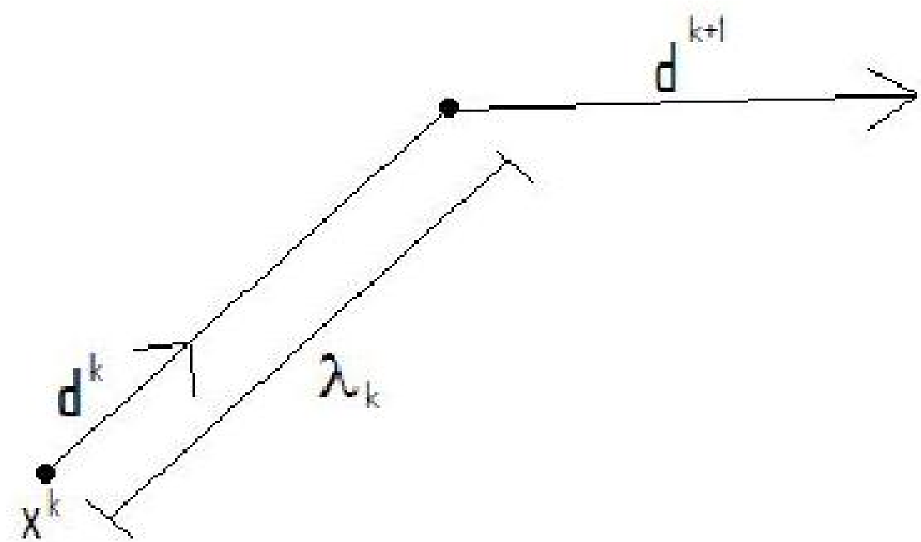
Los métodos de descenso presentados a continuación utilizan esta dirección de descenso para construir el punto x^{k+1} a partir del punto x^k.

Métodos de Descenso para Funciones Diferenciables

El nuevo punto de la sucesión será:

$$x^{k+1} = x^k + \lambda_k d^k, \text{ con } \lambda_k \in R \text{ (paso)}$$

λ_k se determina de modo que $f(x^{k+1}) < f(x^k)$.



Métodos de Descenso para Funciones Diferenciables

Estructura General de un Método de Descenso:

- **Inicialización:** $k = 0$
  Seleccionar convenientemente un punto de partida x^0.

- **Iteración k**
  - Si $\nabla f(x^k) = 0$, entonces x^k es un punto estacionario y se debe finalizar.
  - Si lo anterior no se cumple, se debe determinar una dirección de descenso d^k para f en x^k.
  - Determinar un paso $\lambda_k > 0$ tal que:

$$f(x^{k+1}) = f(x^k + \lambda_k d^k) < f(x^k)$$

  - Definir $x^{k+1} = x^k + \lambda_k d^k$ e incrementar k en 1 para volver a iterar.

Nota: Se pueden colocar criterios de detención alternativos.

Métodos de Descenso para Funciones Diferenciables

Dada la arbitrariedad de la elección de algunos parámetros, se impondrán reglas para su elección, estos parámetros son:

- dirección de descenso d^k
- paso λ_k
- punto de partida x^0

La determinación de λ_k se realiza generalmente mediante algún método de minimización unidimensional:

$$\begin{aligned} &\min_{\lambda} h_k(\lambda) \\ &\text{s.a. } \lambda \geq 0 \end{aligned}$$

Donde $h_k(\lambda) = f(x^k + \lambda d^k)$

Si la función $f(x)$ es convexa, entonces la función $f(x + \lambda d)$ también es convexa para todo x y d fijos. Así, se puede determinar λ_k resolviendo la ecuación $h'_k(\lambda) = 0$.

Métodos de Descenso para Funciones Diferenciables

Luego, utilizando la regla de la cadena obtenemos:

$$\frac{d}{d\lambda}f(x^k + \lambda d^k) = \nabla f(x^k + \lambda d^k)^T d^k$$

Convergencia de los Métodos:

Teorema:

Sea $f : R^n \longrightarrow R$ una función diferenciable y sea $\{x^k\}$ la sucesión generada por el método anterior. Si $\{x^k\}$ está contenida en un conjunto compacto (cerrado y acotado), entonces ocurre una de las siguientes opciones:

- La sucesión es finita y el último punto es estacionario.
- La sucesión es infinita y todo punto de acumulación x^* al que converge la sucesión es estacionario.

Nota: Observar que este resultado es general y no entrega un criterio práctico para asegurar convergencia.

Métodos de Descenso Particulares: Método del Gradiente

Desarrollado por el matemático frances Cauchy, el método calcula en cada punto x^k el gradiente y determina x^{k+1} minimizando f en la dirección de $-\nabla f(x^k)$ desde x^k.

Luego el método viene dado por el paso iterativo:

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k)$$

Con λ_k solución óptima del problema:

$$\min_{\lambda} h_k(\lambda) = f(x^k + \lambda d^k)$$

$$\text{s.a. } \lambda \geq 0$$

Métodos de Descenso Particulares: Método del Gradiente

Ej: $\min f(x,y) = (x-2)^2 + 2x + y^2 - y + 3$

$$\nabla f(x,y) = (2x-2, 2y-1)$$

Tomemos $x^0 = (x,y) = (1,1)$

- **Iteración** $k = 0$
  $\nabla f(1,1) = (0,1) \neq (0,0)$ por lo tanto x^0 no es estacionario.

  $$d^0 = -\nabla f(x^0) = -\nabla f(1,1) = (0,-1)$$

  Luego:

  $$h_1(\lambda) = x^0 - \lambda \nabla f(x^0) = f(1, 1-\lambda)$$
  $$h_1(\lambda) = (1-2)^2 + 2 + (1-\lambda)^2 - (1-\lambda) + 3$$
  $$h_1(\lambda) = \lambda^2 - \lambda + 6$$

Métodos de Descenso Particulares: Método del Gradiente

Así, para encontrar λ_0 se debe resolver el problema:

$$\min_{\lambda} \lambda^2 - \lambda + 6$$

$$\text{s.a. } \lambda \geq 0$$

Como $h_1(\lambda)$ es convexa, el problema equivale a resolver:

$$h_1'(\lambda) = 2\lambda - 1 = 0$$

$$\Rightarrow \lambda_0 = \frac{1}{2}$$

Con esto:

$$x^1 = x^0 + \lambda_0 d^0$$

$$x^1 = (1,1) + \frac{1}{2}(0,-1)$$

$$x^1 = (1, \frac{1}{2})$$

Métodos de Descenso Particulares: Método del Gradiente

- **Iteración** $k = 1$
  $\nabla f(1, \frac{1}{2}) = (0, 0) \Rightarrow x^1$ es punto estacionario de f.

  Es mínimo?
  Veamos el hessiano de f:
  $$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

  Luego $Hf(x, y)$ es definida positiva $\Rightarrow f$ es convexa $\Rightarrow x^1 = (1, \frac{1}{2})$ es mínimo global de f.

En este caso se llega al óptimo en un paso, pero en general, el método del gradiente produce una sucesión infinita de puntos y es necesario establecer criterios de detención apropiados.

Métodos de Descenso Particulares: Método del Gradiente

Posibles Criterios de Detención: (en lugar de $\nabla f(x^k) = 0$)

1. $\|\nabla f(x^k)\| \leq \varepsilon \text{ con } \varepsilon \text{ pequeño}$
2. $|\frac{\partial f}{\partial x_i}| \leq \varepsilon, \forall i = 1, \ldots, n$ con ε pequeño.
3. $\|x^{k+1} - x^k\| \leq \varepsilon$ durante T iteraciones sucesivas.

En aplicaciones prácticas se suele combinar 1 o 2 con 3.

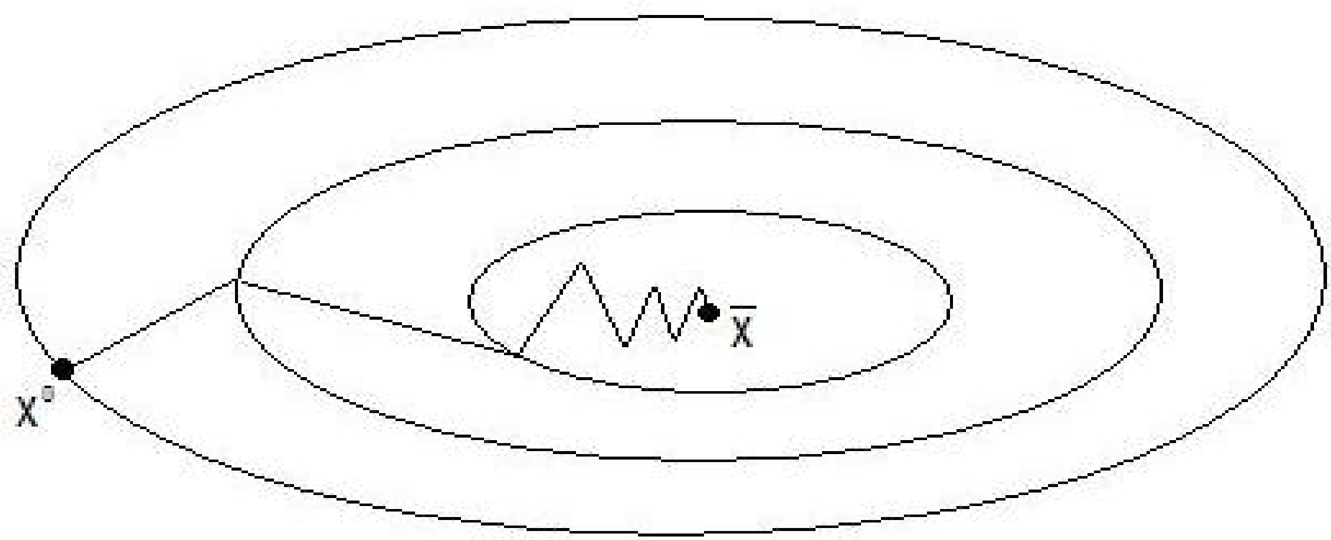
Convergencia del Método del Gradiente:
Si f es continuamente diferenciable y la sucesión $\{x^k\}$ generada por el método está contenida en un conjunto compacto, entonces el método converge a un punto estacionario. Si además f es convexa y el método converge entonces la sucesión generada converge a un punto mínimo de f.
Una condición para que el método converja es que la función f sea tal que $f(x) \longrightarrow +\infty$ cuando $\|x\| \longrightarrow \infty$, o sea, f es acotada inferiormente.

Métodos de Descenso Particulares: Método del Gradiente

Cuando en iteraciones sucesivas del método los gradientes son ortogonales entre sí, la convergencia se vuelve muy lenta para algunos tipos de funciones.

Si las superficies de nivel son hiperesferas (circunsferencias en 2 o más dimensiones), el método encuentra el mínimo en una iteración (como en el ejemplo).

Sin embargo, si las superficies de nivel son excéntricas (elipsoides), la convergencia puede ser muy lenta debido a las oscilaciones.



Métodos de Descenso Particulares: Método de Newton

Busca superar algunas de las dificultades que presenta el método del gradiente. Se basa en aproximar la función f por una función cuadrática (en cada punto x^k se aproxima por la expansión de Taylor de orden 2) y esta aproximación se minimiza exactamente, generando un nuevo punto x^{k+1}.

$$q(x) = f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T Hf(x^k)(x - x^k)$$

Una condición necesaria para un mínimo de $q(x)$ es que x sea punto estacionario de q. Si q es convexa, esta condición es necesaria y suficiente.

Así la condición se traduce en:

$$\nabla q(x) = \nabla f(x^k) + Hf(x^k)(x - x^k) = 0$$

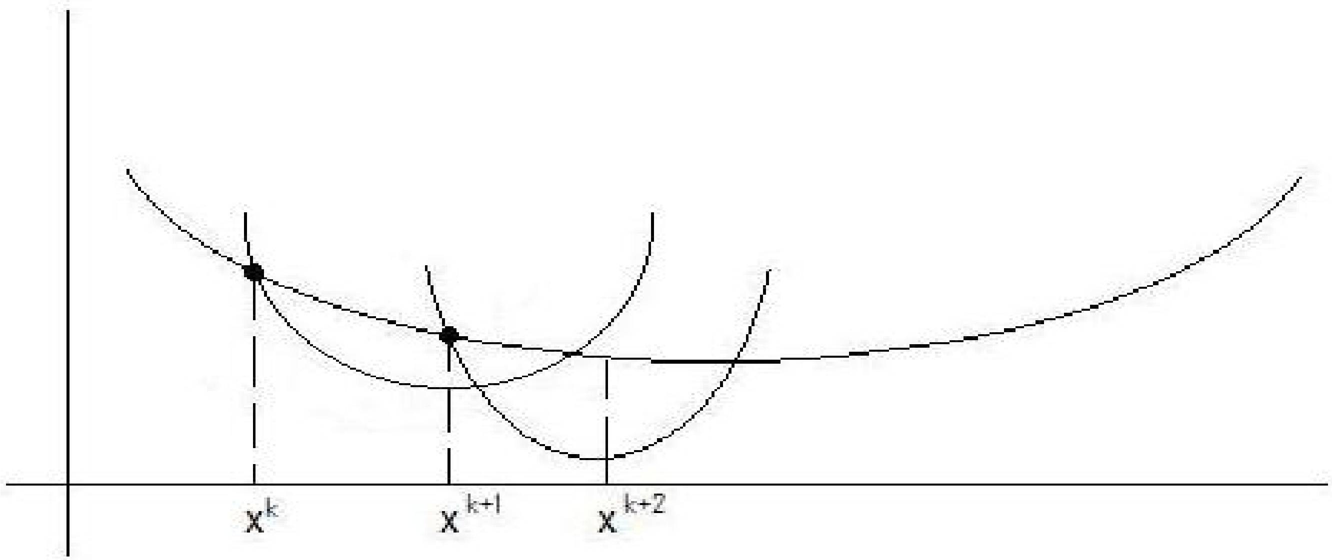
$$\Rightarrow Hf(x^k)(x - x^k) = -\nabla f(x^k)$$

Métodos de Descenso Particulares: Método de Newton

La solución de este problema será el punto x^{k+1}.
Si $Hf(x^k)$ es invertible:

$$\Rightarrow x^{k+1} = x^k - [Hf(x^k)]^{-1}\nabla f(x^k)$$

Se puede visualizar este método como aquel cuya dirección de movimiento en cada x^k es $d^k = -[Hf(x^k)]^{-1}\nabla f(x^k)$ con $\lambda = 1$



Métodos de Descenso Particulares: Método de Newton

Ej:

1. $\min f(x,y) = (x-2)^2 + 2x + y^2 - y + 3$

$$\nabla f(x,y) = (2x-2, 2y-1)$$

El hessiano de f es:

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

La inversa corresponde a:

$$\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

Apliquemos Newton:

- $k = 0$

$$x^0 = (1,1) \text{ y } \nabla f(1,1) = (0,1)$$
$$\Rightarrow x^1 = (1,1) - [Hf(x)]^{-1}(0,1) = (1, \frac{1}{2})$$

Métodos de Descenso Particulares: Método de Newton

- $k=1$
  $\nabla f(x^1)=(0,0)$, luego x^1 es estacionario. Se puede ver que es mínimo pues f es convexa.

2. $\min f(x)=x+e^{-3x}$

$$\nabla f(x)=1-3e^{-3x}$$
$$Hf(x)=9e^{-3x}$$

Luego:

$$x^{k+1}=x^k-\frac{e^{3x^k}}{9}(1-3e^{-3x^k})$$
$$x^{k+1}=x^k-\frac{e^{3x^k}}{9}+\frac{1}{3}$$
$$x^{k+1}=x^k+\frac{1}{3}(1-\frac{1}{3}e^{3x^k})$$

La sucesión de puntos sería:
$x^0=0$
$x^1=0,2222222$

Métodos de Descenso Particulares: Método de Newton

$x^2 = 0,3391406$
$x^3 = 0,3651345$
$x^4 = 0,3662024$
En pocas iteraciones llegamos a un valor muy cercano al óptimo $x^* = \ln(\sqrt[3]{3})$

Convergencia del Método:
La sucesión puede no ser convergente. Por ejemplo, puede pasar que $Hf(x^k)$ sea singular y por consiguiente no invertible. También puede pasar que no se elija convenientemente el punto x^0 y la sucesión converja a un máximo (se puede verificar con el ejemplo $f(x) = x^4 - 24x^2$ con $x^0 = 1$).

Gradiente vs Newton:
Método del Gradiente $\longrightarrow$ Comportamiento lento a medida que se acerca al óptimo.

Método de Newton $\longrightarrow$ Converge más rápidamente pero requiere un punto inicial cercano al óptimo, requiere de invertir una matriz o resolver un sistema de ecuaciones.

Métodos de Descenso Particulares: Método de Newton *Globalizado*

Se agrega una minimización unidimensional en la dirección generada:

$$d^k = -[Hf(x^k)]^{-1}\nabla f(x^k)$$

d^k es dirección de descenso si $Hf(x)$ es definida positiva (y esto pasa si f es convexa).

En cada iteración se puede buscar un paso $\lambda_k > 0$ tal que $f(x^k + \lambda_k d^k) < f(x^k)$ (recordar que Newton trabaja con $\lambda_k = 1$):

$$\text{mín}\, f(x^k + \lambda_k d^k)$$

$$\text{s.a. } \lambda_k \geq 0$$

Bajo ciertas condiciones esta variante garantiza la convergencia a un punto estacionario cualquiera sea el x^0 inicial.

Optimización con Restricciones

$$\begin{array}{ll} \text{(P)} & \min f(x) \\ \text{s.a.} & g_1(x) \leq 0 \\ & \vdots \\ & g_m(x) \leq 0 \end{array}$$

Con f y g_i continuas y diferenciables.
$S = \{x \in R^n / g_i(x) \leq 0, i = 1, \ldots, m\}$ es el conjunto de soluciones factibles de (P).

Observación: No se pierde generalidad planteando el problema (P).

- Se estudiarán condiciones necesarias y suficientes que se satisfacen en los puntos óptimos.
- El análisis es más complejo que en el caso irrestricto pero la idea es similar: se buscan condiciones analíticas que indiquen si la función crece o decrece en ciertas direcciones.
- Luego se verán las ideas en que se basan algunos métodos para resolver este problema.

Caracterización de Puntos Óptimos

Definición:
Sea $x \in S$, un vector $d \in R^n$ es una **dirección factible** en S con respecto a x si existe $\varepsilon > 0$ tal que $x + \lambda d \in S$, $\forall \lambda \in (0, \varepsilon]$

Definición: $D(x) = \{d \in R^n / d$ es dirección factible en S con respecto a $x\}$ es el **conjunto de direcciones factibles**.

Ej:

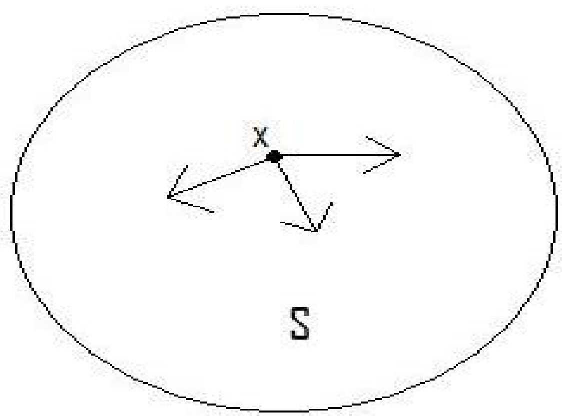


Figura 1: $D(x) = R^n$

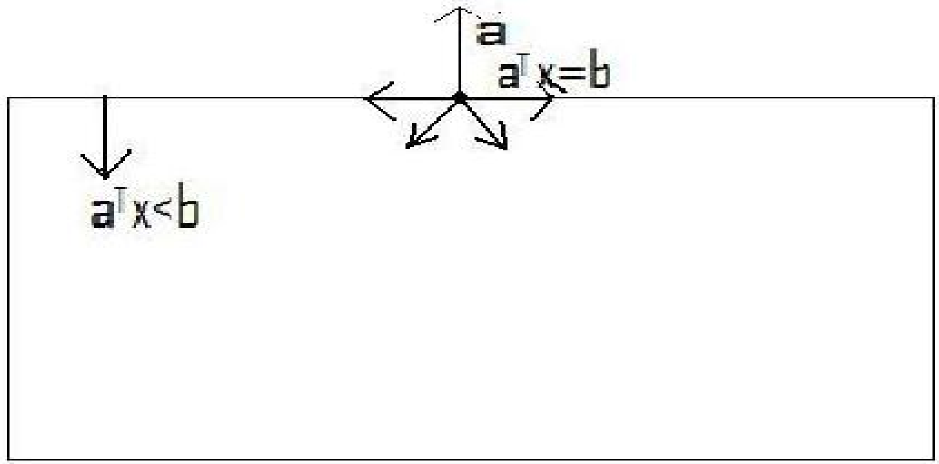


Figura 2: $D(x) = \{d \in R^n / a^T d \leq 0\}$

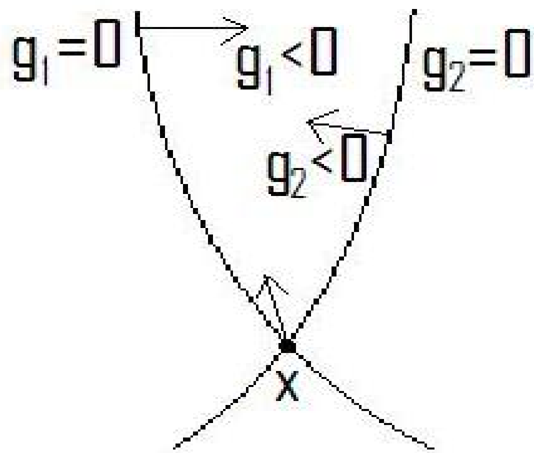


Figura 3: $D(x) = \{d \in R^n / \nabla g_1(x)^T d < 0 \text{ y } \nabla g_2(x)^T d < 0\}$

Caracterización de Puntos Óptimos

Teorema: (Primera Condición de Optimalidad)
Sea $\overline{x}$ un mínimo local del problema (P). Entonces $\nabla f(\overline{x}) \cdot d \geq 0$, $\forall d \in D(\overline{x})$

Demostración:
Sean $d \in D(\overline{x})$ y $\varepsilon > 0$ tales que $\overline{x} + \lambda d \in S$, $\forall \lambda \in (0, \varepsilon]$.
Como $\overline{x}$ es mínimo local, existe $\delta > 0$ tal que $f(x) \geq f(\overline{x})$, $\forall x \in B(\overline{x}, \delta) \bigcap S$.
Por otro lado, si $\nabla f(\overline{x})^T d < 0$, entonces existe $0 < \overline{\varepsilon} < \min\{\varepsilon, \delta\}$ tal que $f(\overline{x} + \lambda d) < f(\overline{x})$, $\forall \lambda \in (0, \varepsilon]$.
Pero $\overline{x} + \lambda d \in S$ y esto contradice el hecho de que $\overline{x}$ es mínimo local de (P).

Observación: Geométricamente, si $\overline{x}$ es mínimo local de (P) $\Rightarrow$ todas las direcciones factibles en $\overline{x}$ tienen ángulo $\leq 90^{\text{o}}$ con el vector $\nabla f(\overline{x})$ (lo que significa que f crece en cualquier dirección factible en $\overline{x}$).
Este resultado no es práctico puesto que no se puede verificar la condición para todas las direcciones factibles en un punto.
Desarrollaremos una condición analítica con el mismo propósito.

Regularidad

Concepto de Regularidad:

Definición:

Una restricción de desigualdad $g_i(x) \leq 0$ es **activa** en un punto factible $\overline{x}$ si $g_i(\overline{x}) = 0$ y es **no activa** si $g_i(\overline{x}) < 0$. (Una restricción $h_j(x) = 0$ es activa en todo punto factible)

Definición:

Sea $x \in S$, el **conjunto activo** es:

$$I(x) = \{i \in \{1, \dots, m\} / g_i(x) = 0\}$$

Definición:

Sea $x \in S$ e $I(x)$ el conjunto activo. Se define el **cono tangente en** x como:

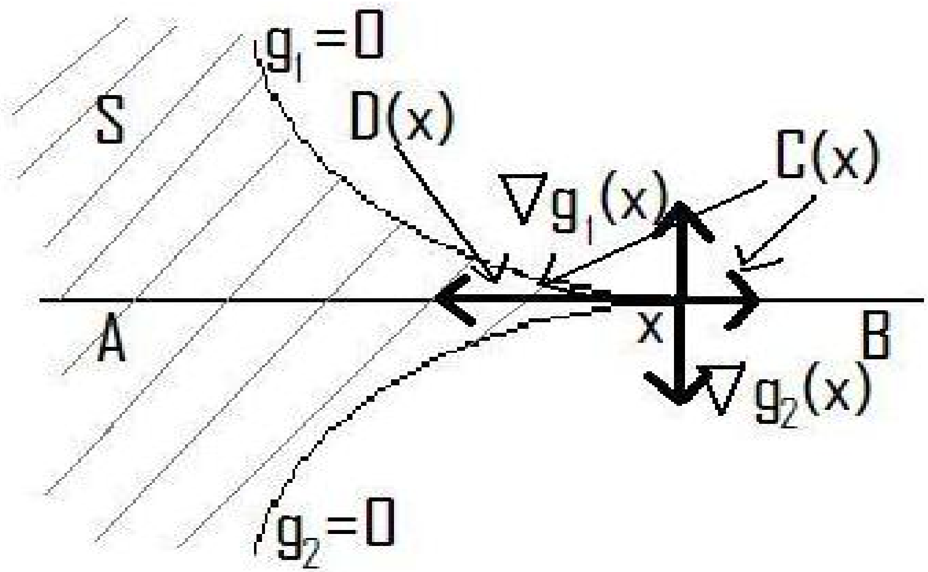
$$C(x) = \{d \in R^n / \nabla g_i(x)^T d \leq 0, \forall i \in I(x)\}$$

Teorema:

$$D(x) \subseteq C(x), \; \forall x \text{ punto factible de (P)}$$

Regularidad

Ej: (donde $D(x) \neq C(x)$)



Regularidad

Definición de Regularidad:
Sea $x \in S$ e $I(x)$ el conjunto activo tal que $I(x) \neq \phi$. Se dice que las funciones g_i ($i \in I(x)$) cumplen la condición de regularidad en x si $C(x) = cl(D(x))$, con cl la clausura del conjunto.

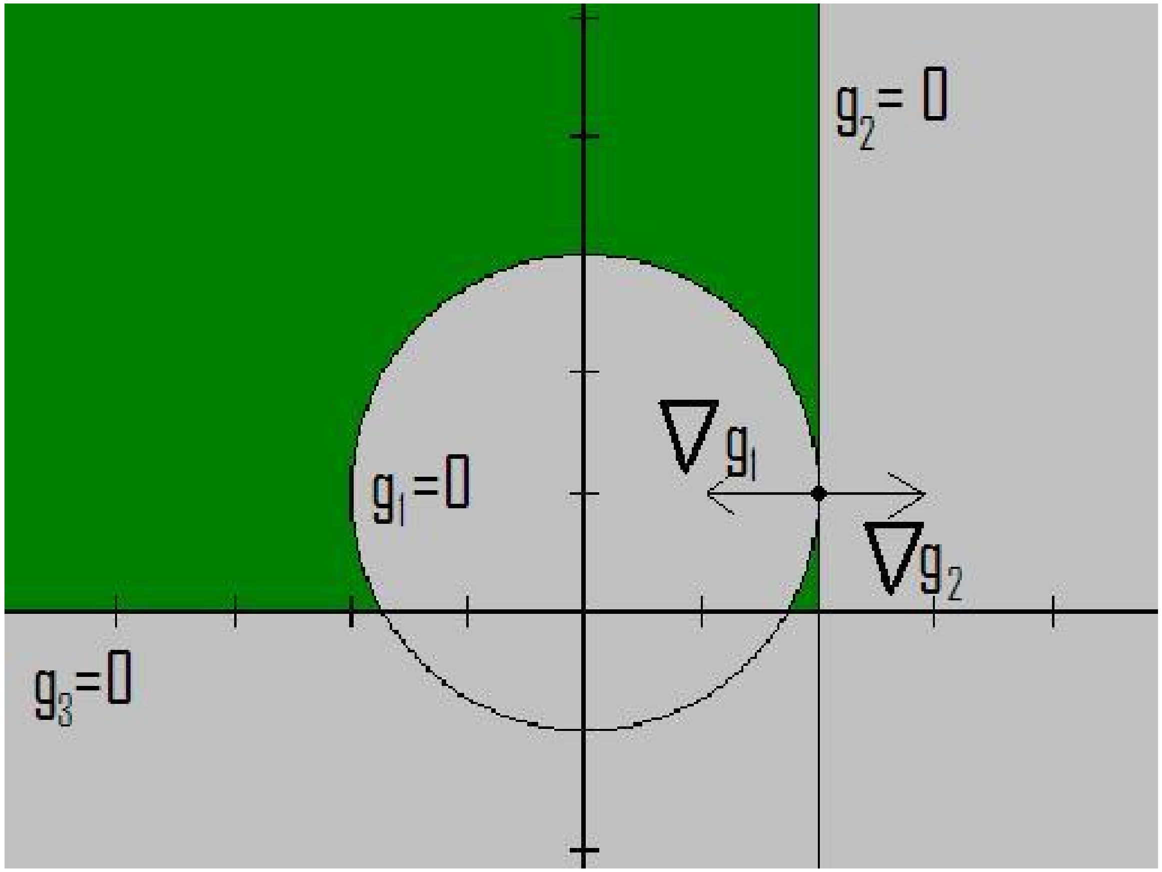
Para funciones diferenciables tenemos la siguiente propiedad:

Propiedad:
x es regular para las restricciones de (P) si los vectores gradientes de las restricciones activas en x son linealmente independientes. La recíproca no es necesariamente cierta.

Ej:

$$g_1 : -x^2 - (y-1)^2 + 4 \leq 0$$
$$g_2 : x - 2 \leq 0$$
$$g_3 : -y \leq 0$$

Regularidad



$(x, y) = (2, 1)$ NO cumple la condición suficiente para ser regular.
Pero $C(x, y) = cl(D(x, y))$
g_1 y g_2 restricciones activas, $\nabla g_1(2, 1) = (-4, 0)^T$ y $\nabla g_2(2, 1) = (1, 0)^T$ son linealmente dependientes.

Caracterización de Puntos Óptimos

Teorema: (Segunda Condición de Optimalidad)
Si $\overline{x}$ es un mínimo local de (P) y se cumple la condición de regularidad en $\overline{x}$, entonces:

$$\nabla f(\overline{x})^T d \geq 0, \forall d \text{ tal que } \nabla g_i(\overline{x})^T d \leq 0, i \in I(\overline{x}) \ (d \in C(\overline{x}))$$

Observación: Aún se tiene una condición no verificable en la práctica. Para traducir esto, usamos un resultado de algebra lineal que relaciona sistemas de desigualdades lineales.

Lema de Farkas:
Sean $A \in R^{m \times n}$ y $b \in R^m$, entonces uno y sólo uno de los siguientes sistemas tiene solución:

- Sistema 1: $Ax = b, x \geq 0$
- Sistema 2: $\mu^T A \geq 0, \mu^T b < 0$

Caracterización de Puntos Óptimos

En otras palabras:

$$\nexists x \in R^n \text{ tal que } Ax = b, x \geq 0 \Leftrightarrow \exists \mu \in R^m \text{ tal que } \mu^T A \geq 0, \mu^T b < 0$$

Así A tiene la forma:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Con esto $\overrightarrow{b}$ queda expresado como:

$$\overrightarrow{b} = x_1 \overrightarrow{A_1} + \cdots + x_n \overrightarrow{A_n}$$

Luego si existe $x \geq 0$ tal que $Ax = b$, entonces b es combinación lineal no negativa de los vectores columna de la matriz A, denotados $A_1, \ldots, A_n$.
Por lo tanto, b está contenido en el cono generado por las columnas de A (y sale fácil que $\nexists \mu$ que resuelva el sistema 2).

Caracterización de Puntos Óptimos

Por otro lado, si **no** existe $x \geq 0$ tal que $Ax = b$, entonces b no está contenido en el cono generado por las columnas de A.
Dado que este cono es convexo, existe un hiperplano H que lo separa del vector b.
Si μ es el vector normal de este hiperplano, esto equivale a que $\mu^T A \geq 0$ (ángulo agudo) y $\mu^T b < 0$ (ángulo obtuso), lo cual corresponde al sistema 2.

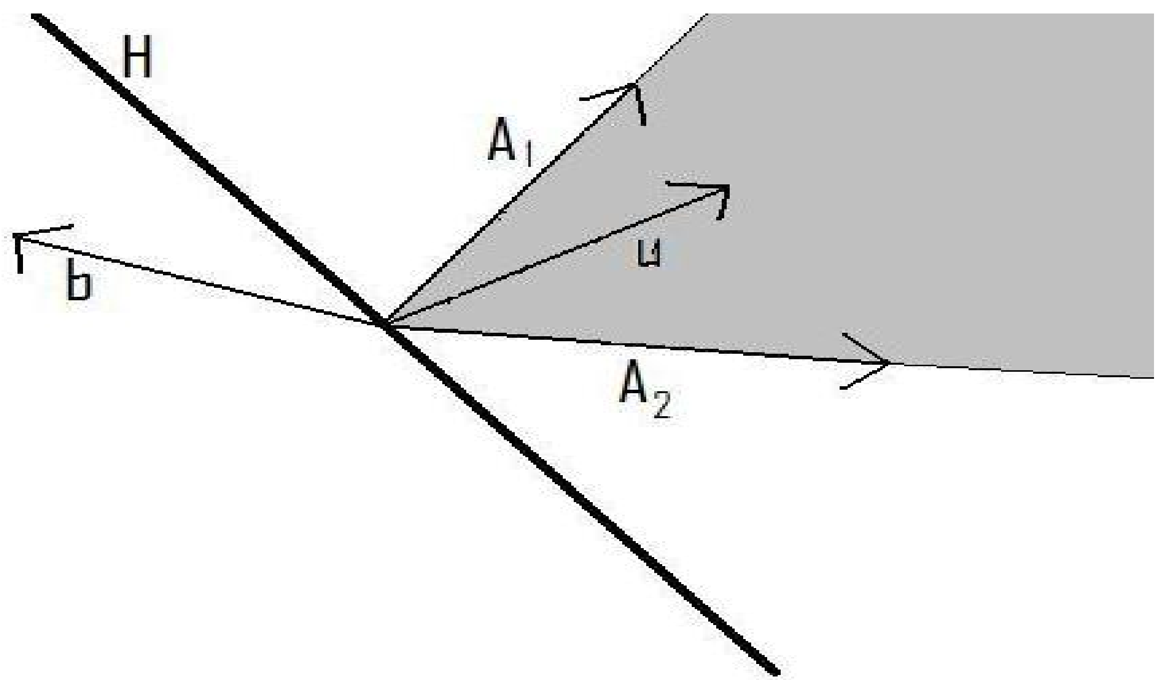


Figura 4: sistema 2

Condición de Karush - Kuhn - Tucker

Teorema: (Condición Necesaria de KKT)
Sea $\overline{x}$ mínimo local tal que se cumple la condición de regularidad en $\overline{x}$. Entonces, existen escalares $\mu_1 \geq 0, \ldots, \mu_m \geq 0$ denominados "multiplicadores de Lagrange", tales que:

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} \mu_i \nabla g_i(\overline{x}) = 0$$

$$\mu_i g_i(\overline{x}) = 0,\ i = 1, \ldots, m$$

La última ecuación se conoce como "Condiciones de Holgura Complementaria".

Demostración:
Si $\overline{x}$ es mínimo local, entonces no existe $d \in R^n$ tal que:

$$\nabla g_i(\overline{x})^T d \leq 0, i \in I(\overline{x}) \text{ y } \nabla f(\overline{x})^T d < 0$$

$$\Leftrightarrow d^T(-\nabla g_i(\overline{x})) \geq 0, i \in I(\overline{x}) \text{ y } d^T \nabla f(\overline{x}) < 0$$

Condición de Karush - Kuhn - Tucker

Así $b = \nabla f(\overline{x})$ y $A_i = (-\nabla g_i(\overline{x})), i \in I(\overline{x})$.

Luego, por el Lema de Farkas:
Existen $\mu_i \geq 0$, $i \in I(\overline{x})$ tal que:

$$\nabla f(\overline{x}) = - \sum_{i \in I(\overline{x})} \mu_i \nabla g_i(\overline{x})$$

Agrego $\mu_i = 0$, $\forall i$ no contenido en $I(\overline{x})$ y tengo:

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} \mu_i \nabla g_i(\overline{x}) = 0$$

Por otro lado:

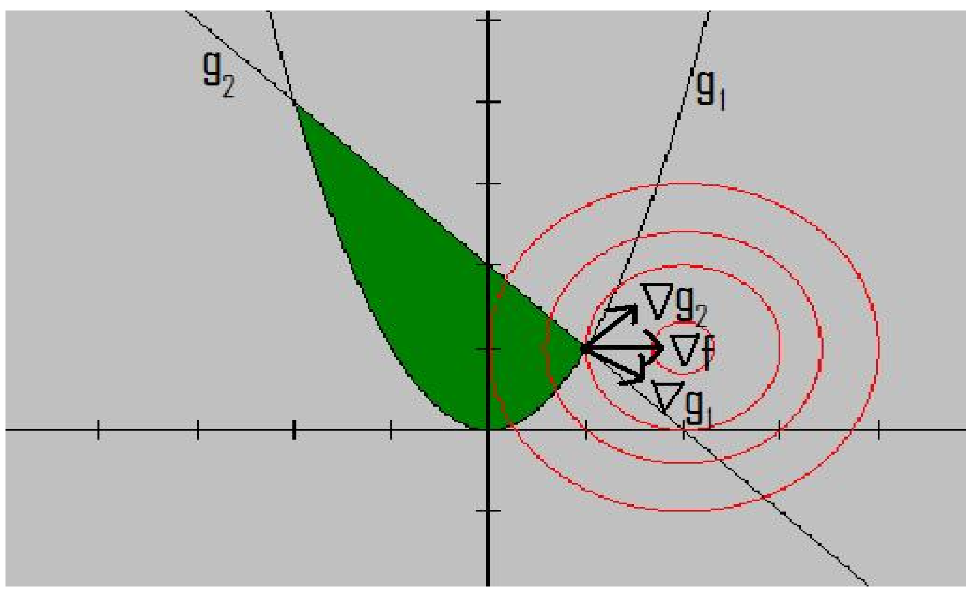
$$g_i(\overline{x}) = 0 \text{ para } i \in I(\overline{x})$$

Condición de Karush - Kuhn - Tucker

$$\mu_i = 0 \text{ para } i \text{ no contenido en } I(\overline{x})$$
$$\Rightarrow \mu_i g_i(\overline{x}) = 0, \ \forall i = 1, \ldots, m$$

Ej:

$$\begin{aligned} \min f(x,y) &= (x-2)^2 + (y-1)^2 \\ s.a. \quad g_1(x,y) &: -y + x^2 \leq 0 \\ g_2(x,y) &: x + y - 2 \leq 0 \end{aligned}$$



Condición de Karush - Kuhn - Tucker

Gráficamente se ve que $\overline{x} = (1,1)$ es óptimo de la función.

En (1,1) las 2 restricciones son activas y se tiene que $\nabla g_1(1,1) = (2,-1)^T$ y $\nabla g_2(1,1) = (1,1)^T$ que son linealmente independientes, luego $\overline{x}$ es un punto regular.

El conjunto de direcciones factibles D en (1,1) son los vectores contenidos entre la recta $x + y - 2 = 0$ y la tangente a la curva $x^2 - y = 0$ en (1,1) $(y = 2x - 1)$.

$-\nabla f$ apunta en la dirección de máximo decrecimiento de f, por lo tanto un pequeño desplazamiento en una dirección que forme un ángulo $< 90^{\text{o}}$ con $-\nabla f$ hará decrecer f. Luego en el óptimo ninguna dirección factible forma ángulo $< 90^{\text{o}}$ con $-\nabla f$.

Si en el (1,1) $-\nabla f$ no estuviese contenido en el cono formado por ∇g_1 y ∇g_2 entonces **existiría** una dirección de descenso de f factible (en el óptimo las direcciones en las cuales la función f decrece son infactibles).

Condición de Karush - Kuhn - Tucker

Veamos las condiciones de KKT:

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} \mu_i \nabla g_i(\overline{x}) = 0$$

$$(2(x-2), 2(y-1))^T + \mu_1(2,-1)^T + \mu_2(1,1)^T = (0,0)$$
$$(-2,0) + (2\mu_1, -\mu_1) + (\mu_2, \mu_2) = (0,0)$$

Luego tenemos el sistema:

$$2\mu_1 + \mu_2 = 2$$
$$-\mu_1 + \mu_2 = 0$$
$$\Rightarrow \mu_1 = \frac{2}{3} \text{ y } \mu_2 = \frac{2}{3}$$

Observación: Notar que la condición es necesaria y además pedimos $\overline{x}$ regular. O sea, podemos encontrar un punto no óptimo que verifique la condición y podemos encontrar un punto óptimo que no la verifique (si el punto no es regular).

Condición de Karush - Kuhn - Tucker

Teorema: (Condición Suficiente de KKT)
Sea el problema (P) con f y $g_i \in C^1$ tal que $f(x)$ y la región factible generada por las g_i, $i = 1, \ldots, m$, son convexas. Sea $\overline{x}$ una solución factible para (P).

Si $\overline{x}$ satisface las condiciones de KKT $\Rightarrow$ $\overline{x}$ es mínimo global de (P)

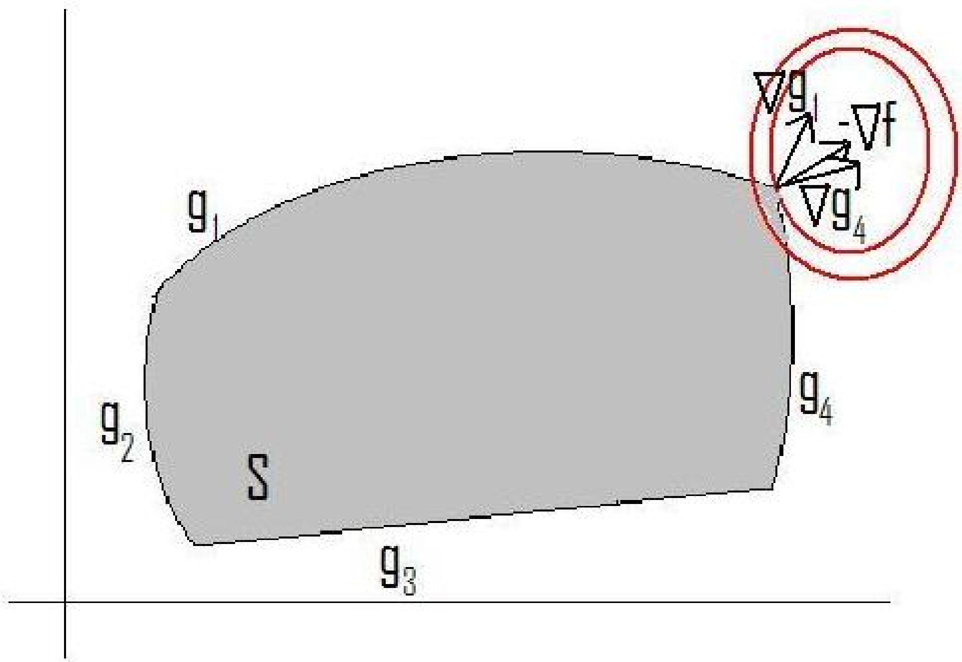
Observación: No se pide regularidad del punto óptimo.

Interpretación Geométrica de KKT:

$$-\nabla f(\overline{x}) = \sum_{i=1}^{m} \mu_i \nabla g_i(\overline{x}) \text{ con , } \mu_i = 0 \text{ para } g_i \text{ restricción no activa}$$

Con f y g_i convexos.
$\Rightarrow -\nabla f(\overline{x})$ puede ser expresado como combinación lineal no negativa de las restricciones activas en $\overline{x}$.
Esto es equivalente a decir que $-\nabla f(\overline{x})$ pertenece al cono generado por los gradientes de las restricciones activas en $\overline{x}$.
Así, cualquier dirección factible forma ángulo obtuso con $-\nabla f(\overline{x})$, o equivalentemente,

la derivada direccional es ≥ 0, esto es, la función f sólo crece localmente en direcciones factibles y como f es convexa tenemos un mínimo global.



Esta condición también es suficiente bajo hipótesis de convexidad.

Ej:
En el ejemplo que se vio recién, las función objetivo y la región factible son convexas y el punto (1,1) pertenece al conjunto y verifica las condiciones.
Así, por el teorema (1,1) es óptimo global del problema.

Condición de Lagrange

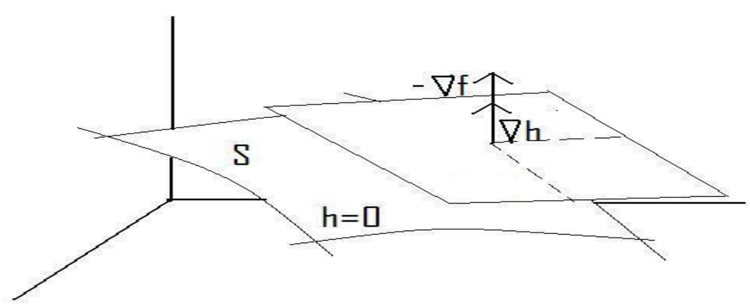
Sea (P_1) un problema de optimización con restricciones sólo de igualdad:

$$\begin{array}{ll} (P_1) & \min f(x) \\ s.a. & h_j(x) = 0, \quad j = 1, \ldots, p \\ & f, h_j \in C^1 \end{array}$$

Teorema: (Condición Necesaria de Lagrange)
Sea $\overline{x}$ un mínimo local de (P_1) tal que $\overline{x}$ es regular. Entonces existen $\lambda_1, \lambda_2, \ldots, \lambda_p \in R$ tal que:

$$\nabla f(\overline{x}) = \sum_{i=1}^{p} \lambda_j \nabla h_j(\overline{x})$$

O sea, ∇f es combinación lineal de los gradientes de las restricciones. (Interpretación Geométrica)



Condición de Lagrange

Demostración:
Las restricciones $h_j(x) = 0$ pueden expresarse como $h_j(x) \leq 0$ y $h_j(x) \geq 0$.
Aplicando la condición necesaria de KKT, tenemos que existen α_j y β_j no negativos tal que:

$$-\nabla f(\overline{x}) = \sum_{i=1}^{p} \alpha_j \nabla h_j(\overline{x}) - \sum_{i=1}^{p} \beta_j \nabla h_j(\overline{x})$$

$$\Rightarrow \nabla f(\overline{x}) = \sum_{i=1}^{p} (\beta_j - \alpha_j) \nabla h_j(\overline{x})$$

Luego, reemplazando $(\beta_j - \alpha_j)$ por λ_j se obtiene el teorema.

Teorema: (Condición Suficiente de Lagrange)
Sea el problema (P$_1$) con $f, h_j \in C^1$, f convexa y las restricciones h_j afines, es decir de la forma $h_j = a_j^T x + b_j$ con $a_j \in R^n$ y $b_j \in R$ y sea $\overline{x}$ un punto factible para (P$_1$). Si $\overline{x}$ satisface la condición de Lagrange, entonces $\overline{x}$ es mínimo global del problema (P$_1$).

El caso general

$$
\begin{array}{lll}
(P_2) & \min f(x) & (f, g_i, h_j \in C^1) \\
s.a. & g_i(x) \leq 0, & i = 1, \ldots, m \\
 & h_j(x) = 0, & j = 1, \ldots, p
\end{array}
$$

La condición de KKT general ahora será la existencia de (x, μ, λ), con x factible, $\mu_i \geq 0$ y $\lambda_j \in R$ tal que:

$$
\nabla f(x) + \sum_{i=1}^{m} \mu_i \nabla g_i(x) + \sum_{j=1}^{p} \lambda_j \nabla h_j(x) = 0
$$

$$
\mu_i g_i(x) = 0, \; i = 1, \ldots, m
$$

Al encontrar un (x, μ, λ) que verifica este sistema (incluyendo la factibilidad), entonces cumple con las condiciones de KKT y es punto factible del problema. Queda sólo por analizar la convexidad para ver si es óptimo.

Un ejemplo

Ej:

1.

$$\begin{array}{lll} \text{mín} & \underbrace{(x-2)^2 + (y-1)^2}_{f(x,y)} & \\ s.a. & \underbrace{x + y - 1}_{h(x,y)} & = 0 \end{array}$$

Luego se tiene el sistema:

$$2(x-2) + \lambda = 0$$
$$2(y-1) + \lambda = 0$$
$$x + y - 1 = 0$$

$$\Rightarrow (x, y, \lambda) = (1, 0, 2)$$

Dado que f es convexa y h es afín $\Rightarrow$ (1,0) es óptimo global del problema.

Otro ejemplo

2. El sistema KKT puede no admitir solución aún si (P_2) tiene óptimos locales. En este caso, ningún mínimo local del problema es regular.
Veamos:

$$\begin{array}{rl} & \min(-y) \\ s.a. & x - 2(1-y)^3 \leq 0 \\ & -x - 2(1-y)^3 \leq 0 \end{array}$$

El sistema KKT es:

$$\mu_1 - \mu_2 = 0$$
$$-1 + 6\mu_1(1-y)^2 + 6\mu_2(1-y)^2 = 0$$
$$x - 2(1-y)^3 \leq 0$$
$$-x - 2(1-y)^3 \leq 0$$
$$\mu_1[x - 2(1-y)^3] = 0$$
$$\mu_2[-x - 2(1-y)^3] = 0$$
$$\mu_1, \mu_2 \geq 0$$

Continuación del ejemplo

Sea $\mu_1 = \mu_2 = \mu$, el multiplicador común.
Luego tenemos que:

$$-1 + 12\mu(1-y)^2 = 0 \Rightarrow \mu \neq 0$$

Dado lo anterior, se tiene:

$$x - 2(1-y)^3 = 0$$

$$-x - 2(1-y)^3 = 0$$

Restando una ecuación con la otra se llega a que $x = 0$.
Luego $-2(1-y)^3 = 0 \Rightarrow (1-y) = 0$ que contradice la ecuación $-1+12\mu(1-y)^2 = 0$
$\Rightarrow$ **no hay solución**

Sin embargo se puede probar que este problema tiene mínimo global en (0,1), pero no hay regularidad en ese punto.