Optimización Continua…Continuación

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. $\|\nabla f(x^k)\| \leq \varepsilon$ con ε 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.

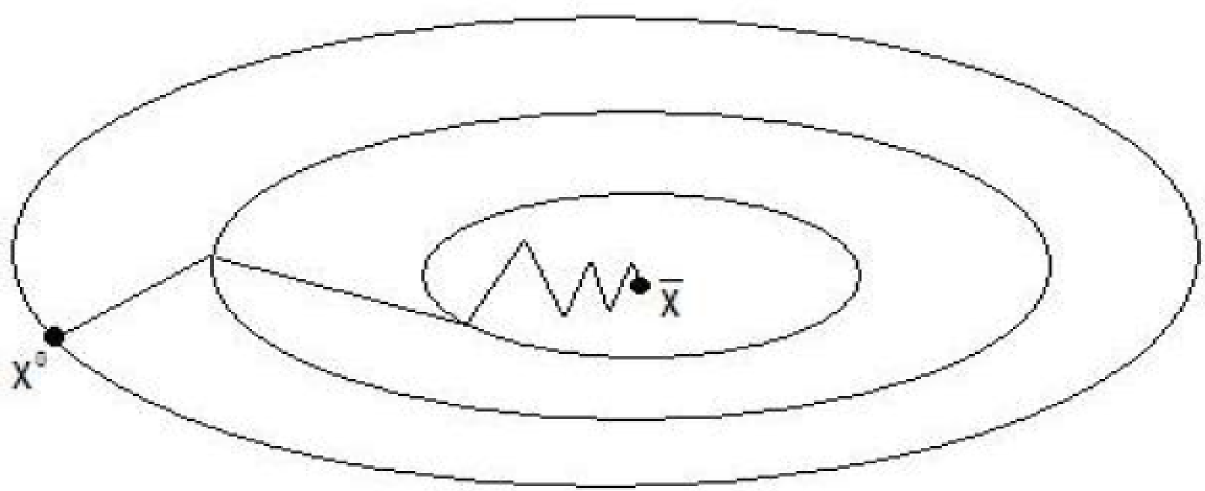
En aplicaciones prácticas se suele combinar 1 o 2 con 3.

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:

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

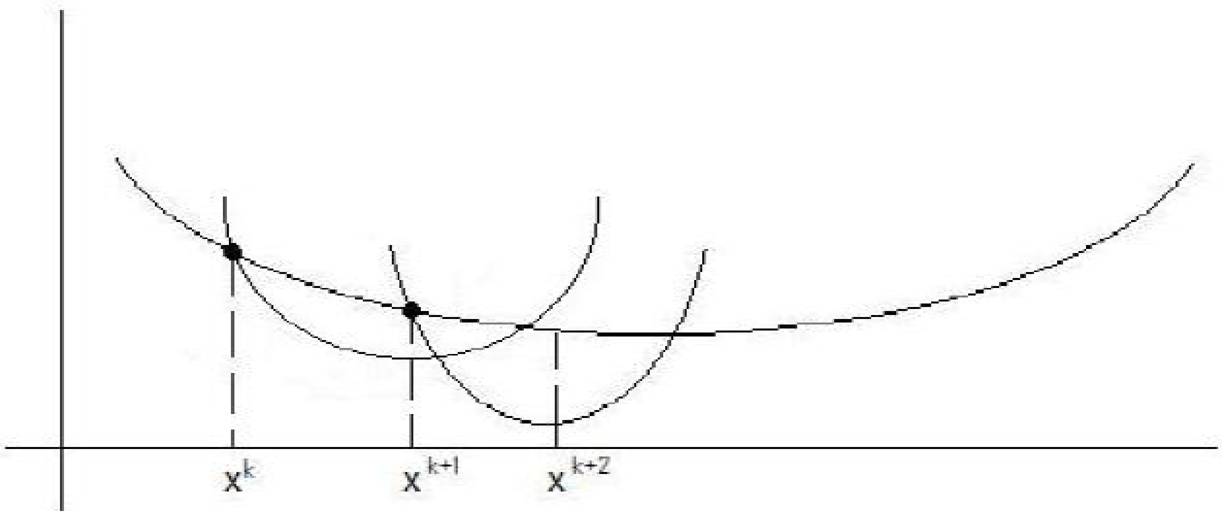
$$\Rightarrow Hf(x^k)(x - x^k) = -\nabla f(x^k)$$

Métodos de Descenso Particulares: Método de Newton

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

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

Se puede visualizar este método como aquel cuya dirección de movimiento en cada x^k es $d^k = -[H\,f(x^k)]^{-1} \nabla f(x^k)$ 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 $x^* = \ln(\sqrt[3]{3})$

Problema bajo dominio restringido:

Consideramos problemas del tipo:

$$(P): \qquad \min f(x) : x \in \Omega \subseteq \mathbb{R}^n$$

donde $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^1(\Omega)$, Ω es un conjunto cerrado. Todo $x \in \Omega$ se dice factible para (P).

Definición Óptimo Global

$x^* \in \Omega$ es un óptimo global para (P) si $f(x^*) \leq f(x)$, $\forall x \in \Omega$.

Definición Óptimo Local

$x^* \in \Omega$ es un óptimo local para (P) si $\exists \varepsilon > 0$ tal que $f(x^*) \leq f(x)$, $\forall x \in \Omega \cap B(x^*, \varepsilon)$.

Definición Problema Convexo

(P) se dice un problema convexo si f es función convexa y si Ω es conjunto convexo.

Teorema de Óptimos Globales para Problemas Convexos

Si (P) es un problema convexo, entonces todos los óptimos locales para (P) son óptimos globales para (P).

- ¿Qué debemos pedir de Ω?

$$\Omega := \{x \in \mathbb{R}^n : g_i(x) \leq 0, \forall i \in I\},$$

donde $g_i : \mathbb{R}^n \to \mathbb{R}, g_i \in \mathcal{C}^1(\mathbb{R}^n)$ y donde $|I| = m \in \mathbb{Z}$.

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**.

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

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:

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

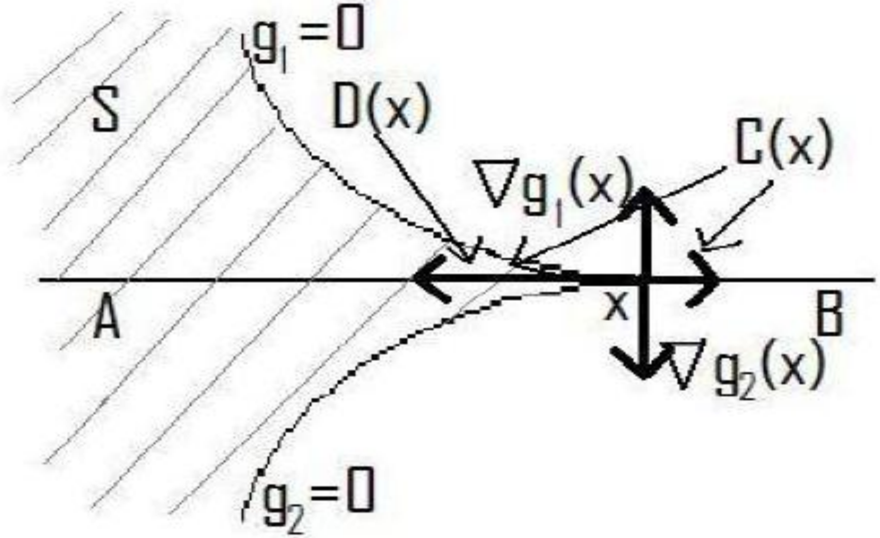
Definición:
Sea $x \in S$ e $I(x)$ el conjunto activo. Se define el **cono tangente en** x como:

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

Teorema:

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

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



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.

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:

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

Condición de Karush - Kuhn - Tucker

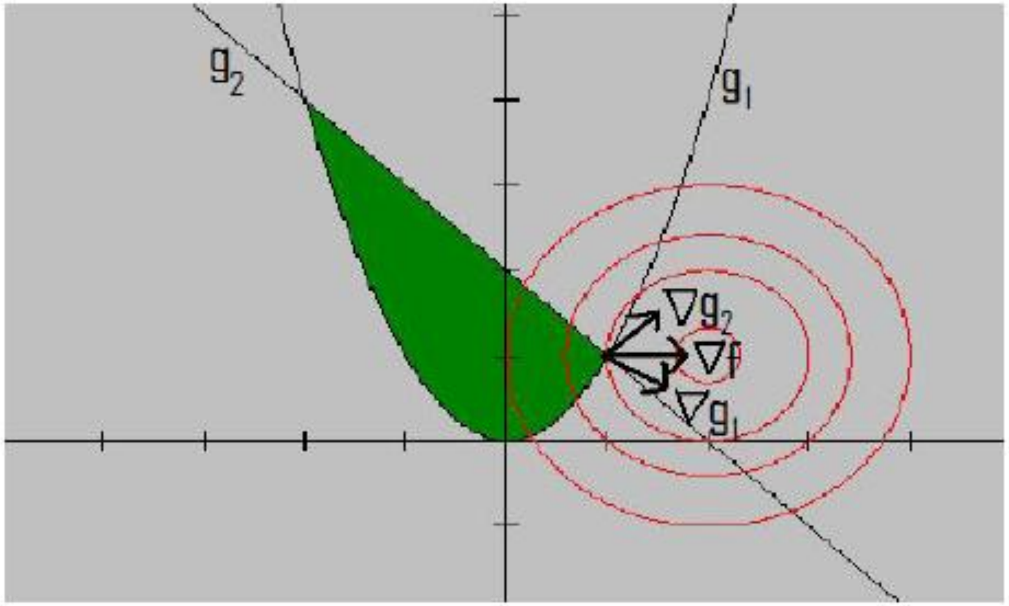
Teorema: (Condición Necesaria de KKT)
Sea $\overline{x}$ mínimo local tal que se cumple la condición de regularidad en $\overline{x}$. Entonces, existen escalares $\mu_1 \geq 0, \ldots, \mu_m \geq 0$ denominados "multiplicadores de Lagrange", tales que:

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

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

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

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



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)

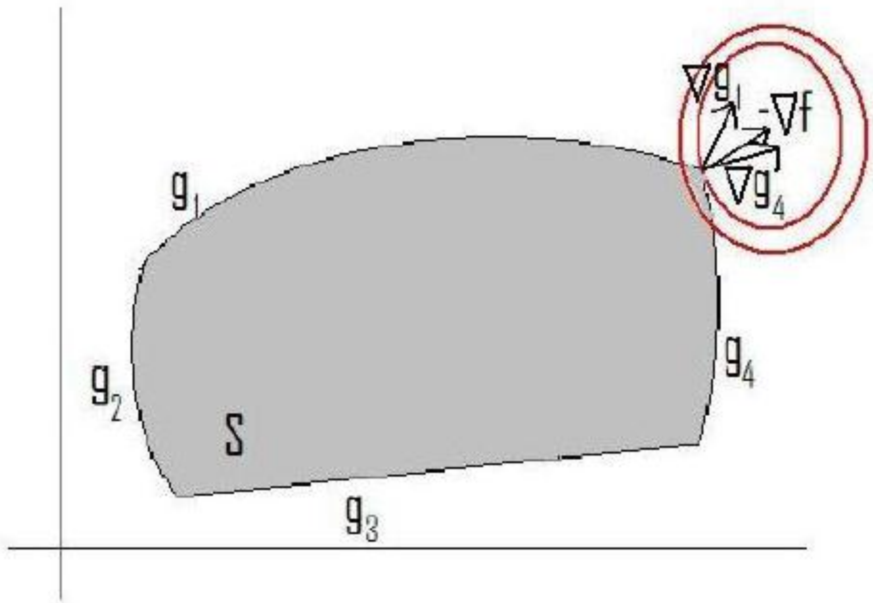
Observación: No se pide regularidad del punto óptimo.

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.

Conjuntos Convexos a partir de Funciones Convexas

Sea $\Omega := \{x \in \mathbb{R}^n : g_i(x) \leq 0,\ i = 1, \ldots, m\}$, si g_i es convexa en $\mathbb{R}^n$, entonces Ω es un conjunto convexo.