

$$\begin{array}{ll} \min J(\theta) & \leftrightarrow \min J(\theta') \\ \text{s.t. } \theta_i > 0 & \theta = e^{\theta'} \end{array}$$

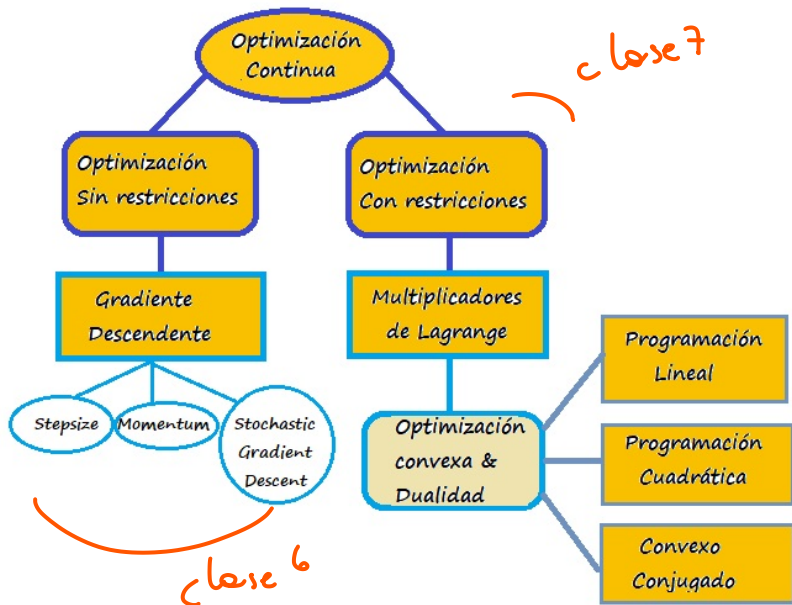
Clase 7: Optimización con restricciones

Verónica Pastor, Martín Errázquin

Análisis Matemático para Inteligencia Artificial

Clase 7

Ramas Principales de la Optimización



Optimización con restricciones

Optimización con restricciones de igualdad

Dado el problema de optimización con restricciones:

$$\begin{aligned} & \text{opt } f(x_1, \dots, x_n) \\ & \text{s.a. } g_1(x_1, \dots, x_n) = 0 \\ & \quad \vdots \\ & \quad g_m(x_1, \dots, x_n) = 0 \end{aligned}$$

$$\begin{aligned} & x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2 = 1 \\ & \|\bar{x}\| = 1 \\ & \underbrace{x_1^2 + \dots + x_n^2 - 1}_{(1) \ g(\bar{x})} = 0 \end{aligned}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ con $i = 1, \dots, m$.

Se denomina **función Lagrangiana** a la función de $n + m$ variables \mathcal{L} definida por:

$$\mathcal{L}(\bar{\lambda}, \bar{x}) = f(\bar{x}) + \lambda_1 g_1(\bar{x}) + \dots + \lambda_m g_m(\bar{x}), \quad \text{con } \bar{\lambda} = (\lambda_1, \dots, \lambda_m)$$

Teorema de condición necesaria de Lagrange

$\# \text{ res.} < \# \text{ vars.}$

Sea $D \subseteq \mathbb{R}^n$ abierto. Sea (1) con $m < n$, donde f, g_i $i = 1, m$ son funciones definidas de D en \mathbb{R} , con derivadas parciales primeras continuas en D y

$$B = \{\bar{x} \in D : g_i(\bar{x}) = 0, i = 1, m\}$$

el conjunto de soluciones factibles.

Entonces, si \bar{x}^* es un óptimo local de (1) tal que la matriz jacobiana de $\bar{g}(\bar{x}^*)$ tal que $|J\bar{g}(\bar{x}^*)_m| \neq 0$, existen m números reales $\lambda_1^*, \dots, \lambda_m^*$ tales que son solución del sistema:

$$\nabla f(\bar{x}^*) - \sum_{i=1}^m \lambda_i^* \cdot \nabla g_i(\bar{x}^*) = \bar{0}$$

$$J_{\bar{g}}(\bar{x}^*) = \begin{bmatrix} -\nabla g_1(\bar{x}^*) \\ \vdots \\ -\nabla g_m(\bar{x}^*) \end{bmatrix}$$

Las soluciones factibles que verifican esta ecuación se denominan *puntos estacionarios*.

Los números $\lambda_1, \dots, \lambda_m$ se denominan **multiplicadores de Lagrange** asociados a las m restricciones en el punto \bar{x}^* .

Ejemplo

$$\min f(x_1, x_2) = x_1^2 + 2x_2 \quad \text{obj}$$

$$\text{s.a. } x_1 + x_2 = 1 \quad \text{restr.}$$

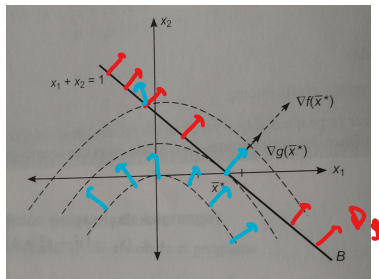
$$g(\bar{x}) = 0$$

↓

$$\nabla g(\bar{x}) = (1, 1)$$

Así definimos:

$$g(x_1, x_2) = x_1 + x_2 - 1.$$



- El conjunto de soluciones factibles es la recta $x_2 = -x_1 + 1$.
- Las curvas de nivel de la función $f(x_1, x_2) = x_1^2 + 2x_2$ son parábolas de la forma $x_1^2 + 2x_2 = k$, $k \in \mathbb{R}$.

$$\therefore \bar{x}^* = (1, 0) \text{ y } f(\bar{x}^*) = 1.$$

Calculamos los vectores gradientes: $\nabla f(x_1, x_2) = (2x_1, 2)$ y $\nabla g(x_1, x_2) = (1, 1)$, ambos vectores son l.d. en \bar{x}^* .

$$\begin{cases} 2x_1 - \lambda \cdot 1 = 0 \\ 2 - \lambda = 0 \end{cases}$$

↓
 $\lambda = 2$

$$\bar{x}^* = (1, 0)$$

Optimización con restricciones de desigualdad

Condiciones necesarias de primer orden de Fritz-John

Dado el problema de optimización con restricciones:

$$\begin{array}{ll} \text{opt } f(x_1, \dots, x_n) & \bar{x} \in \mathbb{R}^n \\ \text{s.a. } g_1(x_1, \dots, x_n) \leq 0 & \\ \vdots & \\ g_m(x_1, \dots, x_n) \leq 0 & \end{array} \quad \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} m \text{ restr.} \quad (2)$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ con $i = 1, \dots, m$. Sea \bar{x}^* un punto tal que $I = \{i : g_i(\bar{x}^*) = 0\}$, f, g_i diferenciables en \bar{x}^* . Entonces, si $g_i, i \notin I$ son continuas en \bar{x}^* se verifica que es solución y existen escalares $\lambda_0, \lambda_i, i \in I$ no todos nulos tales que:

$$\lambda_0 \nabla f(\bar{x}^*) + \sum_{i=1, m} \lambda_i \nabla g_i(\bar{x}^*) = \bar{0}$$

para $0 \leq \lambda_0, \lambda_i, i \in I$, y $g_i(\bar{x}^*) \leq 0, i = 1, m$.

Dualidad en Optimización: Problema dual

Dado el problema primal:

$$\min f(x_1, \dots, x_n)$$

$$\text{s.a. } \left. \begin{array}{l} g(x_1, \dots, x_n) \leq b \\ x_i \geq 0 \end{array} \right\} \text{ m. rest.}$$

$$\bar{x} \in \mathbb{R}^n$$

Tenemos el problema dual:

$$\max D(y_1, \dots, y_m)$$

$$\text{s.a. } \left. \begin{array}{l} h(y_1, \dots, y_m) \geq b \\ y_j \leq 0 \end{array} \right\}$$

$$\bar{y} \in \mathbb{R}^m$$

- Si \bar{x}^* es una solución factible del problema primal e \bar{y}^* es una solución del dual entonces $f(\bar{x}^*) \geq D(\bar{y}^*)$.
- Si un problema no tiene un óptimo finito entonces el otro no es factible.

Función Convexa

Def: Un subconjunto $S \subset \mathbb{R}^n$ es **convexo** si para cada par de puntos $x, y \in S, \alpha \in [0, 1]$ se verifica que $z = \alpha x + (1 - \alpha)y \in S$.

Obs: X_1, X_2 son dos conjuntos convexos, entonces,

- $X_1 \cap X_2$ es convexo.
- $X_1 + X_2 = \{x_1 + x_2 \in \mathbb{R}^n : x_1 \in X_1, x_2 \in X_2\}$.
- si L es una transformación lineal, $L(X_1)$ es convexa.



Función Convexa

Def: Sea $M \subset \mathbb{R}^n, M \neq \emptyset$ convexo, $f : M \rightarrow \mathbb{R}$. Entonces se dice que:

- f es convexa en M sii $\forall x, y \in M, \forall \alpha \in [0, 1]$ se verifica que:

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$



Desigualdad de Jensen

- f es estrictamente convexa en M sii $\forall x, y \in M, \forall \alpha \in (0, 1)$ se verifica que:

$$f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$$

Condiciones para convexidad de funciones diferenciables

Sea $M \subset \mathbb{R}^n$, $M \neq \emptyset$ abierto convexo, $f : M \rightarrow \mathbb{R}$ diferenciable, se dice que:
 f es convexa en $M \Leftrightarrow \forall x, y \in M, f(y) \geq f(x) + \nabla f(x)(y - x)$
O bien, $(\nabla f(y) - \nabla f(x))(y - x) \geq 0$

Prop: Sea $M \subset \mathbb{R}^n$, $M \neq \emptyset$ abierto convexo, $f : M \rightarrow \mathbb{R}$ $f \in \mathcal{C}^2$. Entonces,
 f es convexa en M sii $\forall x \in M, y^T Hf(x)y \geq 0$, para cualquier $y \in \mathbb{R}^n$.

Condiciones Suficientes de Optimalidad Global

$$\begin{aligned} & \text{opt } f(x_1, \dots, x_n) \\ & \text{s.a. } g_1(x_1, \dots, x_n) = 0 \\ & \quad \vdots \\ & \quad g_m(x_1, \dots, x_n) = 0 \end{aligned}$$

con $m < n$ donde f y $g_i, i = 1, \dots, m$ son funciones \mathcal{C}^1 en un conjunto abierto $D \in \mathbb{R}^n$. Entonces se verifica que si f es convexa en el conjunto de soluciones factibles B y, las funciones g_i son lineales, todos los puntos estacionarios son mínimos globales.

Programación Lineal

Un problema de programación lineal es donde tanto la función objetivo como las funciones que definen las restricciones son lineales. La forma general es:

$$\text{opt } c_1x_1 + \dots + c_nx_n$$

$$\text{s.a. } a_{11}x_1 + \dots + a_{1n}x_n \leq b_1$$

$$\vdots$$

$$a_{m1}x_1 + \dots + a_{mn}x_n \leq b_m$$

$$d_{11}x_1 + \dots + d_{1n}x_n \geq e_1$$

$$\vdots$$

$$d_{r1}x_1 + \dots + d_{rn}x_n \geq e_r$$

$$g_{11}x_1 + \dots + g_{1n}x_n = b_1$$

$$\vdots$$

$$g_{s1}x_1 + \dots + g_{sn}x_n = b_s$$

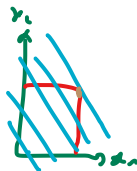
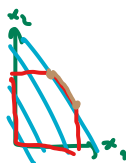
Propiedades de la Programación Lineal

Definición: Sea $S \subset \mathbb{R}^n$ convexo $S \neq \emptyset$, $\bar{x}^* \in S$ es un punto extremo de S si \bar{x}^* no puede expresarse como combinación lineal convexa de puntos de S distintos de él.

- 1 Es un problema convexo ya sea de minimización o maximización.
- 2 La solución óptima, si existe, es global.
- 3 Nunca existen óptimos locales que no sean globales.
- 4 Puede tener o no solución, en caso de existir se encuentra en único punto o en infinitos.

Objetivo: Se busca hallar $\min \bar{c}\bar{x}$, s.a. $A\bar{x} = \bar{b}$; $\bar{x} \geq 0$,
donde $A \in \mathbb{R}^{m \times n}$, $m < n$, $\text{rg}(A) = m$.

$\max z_1 + x_1$



Algoritmo del simplex

Iniciar: eligiendo una solución básica factible de una submatriz B de A , $|B| \neq 0$. *Iterar:*

- ❶ Resolver el sistema $B\bar{x}_B = \bar{b}$. Entonces la solución será $\bar{x} = (\bar{x}_B, \bar{x}_N)$.
- ❷ Calcular $Y = B^{-1}N$, donde $N \in \mathbb{R}^{m \times (n-m)}$ cuyas columnas corresponden a las soluciones no básicas. Hallar $z_j = \bar{c}_B \bar{y}_j$, donde \bar{c}_B son los coeficientes de la función objetivo asociados a la solución básica e \bar{y}_j es la columna de la matriz T correspondiente a x_j .
Sea $z_k - c_k = \max\{z_j - c_j, j \in J\}$ donde J los índices asociados a las variables no básicas.
 - Si $z_k - c_k \leq 0$, la solución básica factible es el óptimo del problema, que no será único si $z_k - c_k = 0$
 - Si $z_k - c_k > 0$ debemos seguir...
- ❸ Antes de construir una nueva solución conviene ver si existe $j \in J : z_k - c_k > 0, \bar{y}_j \leq 0$. Si esto no sucede, terminamos ya que el problema carece de óptimo finito. Sino, construimos una nueva solución:

La variable x_k correspondiente al $z_k - c_k$ máximo, será básica en la nueva solución y dejará de serlo x_l , cuando l sea el índice determinado por:

$$\frac{x_l}{y_{lk}} = \min\left\{\frac{x_i}{y_{ik}} : y_{ik} > 0\right\}$$

Actualizar la nueva submatriz B y volver a (1).

Programación cuadrática

En el caso de que la función objetivo sea una cuadrática convexa:

$$\min_{x \in \mathbb{R}^d} \frac{1}{2} x^T Q x + c^T x \quad \text{y LASSO}$$

$$\text{s.a: } Ax \leq b \quad \text{donde } A \in \mathbb{R}^{m \times d}, b \in \mathbb{R}^m, c \in \mathbb{R}^d.$$

La matriz $Q \in \mathbb{R}^{d \times d}$ es simétrica y definida positiva, y por lo tanto la función objetivo es convexa. El langrangiano está dado por:

$$L(x, \lambda) = \frac{1}{2} x^T Q x + c^T x + \lambda^T (Ax - b)$$

Derivando respecto a x , y despejando:

$$Qx + (c + A^T \lambda) = 0 \Rightarrow x = -Q^{-1}(c + A^T \lambda)$$

Y substituyendo obtenemos la lagrangiana dual:

$$D(\lambda) = -\frac{1}{2}(c + A^T \lambda)^T Q^{-1}(c + A^T \lambda) - \lambda^T b.$$

La **transformada de Legendre-Fenchel** es una transformación de una función convexa diferenciable $f(x)$ a una función que depende de las tangentes $s(x) = \nabla_x f(x)$.

Se define el conjugado convexo de una función $f : \mathbb{R}^D \rightarrow \mathbb{R}$ es una función f^* definida por:

$$f^*(s) = \sup_{x \in \mathbb{R}^D} (\langle x, s \rangle, f(x))$$

Librerías de Python para problemas de optimización

Citado de la página web: <https://relopezbriega.github.io/blog/2017/01/18/problemas-de-optimizacion-con-python/>

En Python podemos encontrar varias librerías que sirven para los problemas de optimización:

- CVXopt: Esta es una librería con una interface para resolver problemas de optimización convexa.
- PuLP: Esta librería proporciona un lenguaje para modelar y resolver problemas de optimización utilizando programación lineal.
- Pyomo: Tiene una notación similar a la que utilizaríamos en la definición matemática de los problemas.

Scipy