Clase 7: Optimización con restricciones

Verónica Pastor, Martín Errázquin

Análisis Matemático para Inteligencia Artificial

Clase 7

Ramas Principales de la Optimización

$f: D \subset \mathbb{R}^n \to \mathbb{R}$

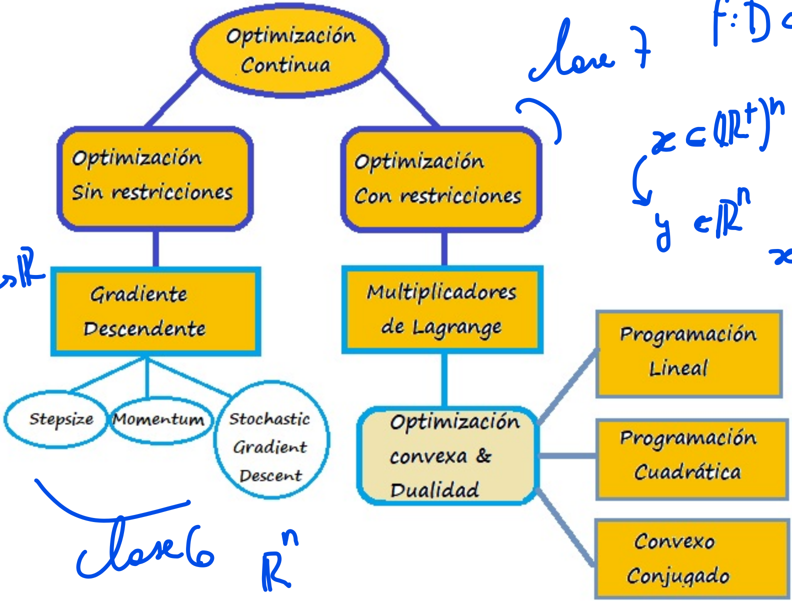
clase 7

$x \in (\mathbb{R}^+)^n$

$y \in \mathbb{R}^n$

$x = e^y$

$f: \mathbb{R}^n \to \mathbb{R}$



clase 6

$\mathbb{R}^n$

Optimización con restricciones

Optimización con restricciones de igualdad

Dado el problema de optimización con restricciones:

$\|x\| = 1$

$x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2 = 1$

$\Updownarrow$

n variables

$$opt\ f(x_1, ..., x_n) \qquad g(x) = x_1^2 + x_2^2 + \ldots + x_n^2 - 1 = 0$$

$$\left.\begin{aligned} \text{s.a. } g_1(x_1, ..., x_n) &= 0 \\ &\vdots \\ g_m(x_1, ..., x_n) &= 0 \end{aligned}\right\} \text{m restricciones} \qquad (1)$$

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

Se denomina función Langrangiana a la función de $n + m$ variables $\mathscr{L}$ definida por:

$\in \mathbb{R}^m$

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

Teorema de condición necesaria de Lagrange

#restr. < #var.

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 $\lambda_1^*, ..., \lambda_m^*$ tales que son solución del sistema:

$\sum \lambda_i^* \cdot \nabla_{g_i}(\bar{x}^*) = -\nabla f(\bar{x}^*)$

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

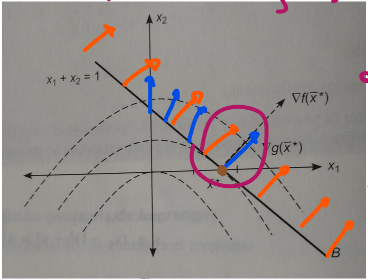
Las soluciones factibles que verifican esta ecuación se denominan *puntos estacionarios*.
Los números $\lambda_1, ..., \lambda_m$ se denominan multiplicadores de Langrange asociados a las m restricciones en el punto $\bar{x^*}$.

$\nabla f(x_1,x_2) = (2x_1, 2)$

$$min\ f(x_1,x_2) = x_1^2 + 2x_2$$

$\forall \bar{x}^*\ \exists \lambda \,/\, \nabla f(\bar{x}^*) = -\lambda \cdot \nabla g(\bar{x}^*)$ ⇒ deben quedar alineados $\nabla f, \nabla g$

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



$\nabla g(x_1,x_2) = (1,1)$

∇g

Así definimos:

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

$\to B$

- 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 $x_1^2 + 2x_2 = k, k \in \mathbb{R}$.

$\left\{ \begin{array}{l} 2x_1 = -\lambda \cdot 1 \\ 2 = -\lambda \cdot 1 \to \lambda = -2 \end{array} \right.$ $\quad x_1 = \frac{-2}{-2} = 1$ $\downarrow$ en B $x_2 = 0$

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

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

Condiciones necesarias de primer orden de Fritz-John

$\|\bar{x}\|^2 \leq 1$

$x_1^2 + x_2^2 + \dots + x_n^2 \leq 1$

Dado el problema de optimización con restricciones:

$g(\bar{x}) = x_1^2 + x_2^2 + \dots + x_n^2 - 1 \leq 0$

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

$\Updownarrow$

$\begin{cases} g_i(\bar{x}) \leq 0 \\ -g_i(\bar{x}) \leq 0 \end{cases}$

$$
\begin{aligned}
& opt\ \ f(x_1, \dots, x_n) \\
& \text{s.a. } g_1(x_1, \dots, x_n) \leqslant 0 \\
& \vdots \\
& g_m(x_1, \dots, x_n) \leqslant 0
\end{aligned}
\tag{2}
$$

donde $f : \mathbb{R}^n \to \mathbb{R}$ y $g_i : \mathbb{R}^n \to \mathbb{R}$ 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:

$-\lambda_0 \nabla f(\bar{x}) = \sum \lambda_i \nabla g(\bar{x})$

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

$(\lambda_0, \lambda_1, \dots, \lambda_m) \neq \vec{0}$

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

Dualidad en Optimización: Problema dual SVM

Dado el problema primal:

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

n vars
m restr.

$$\text{s.a. } \mathrm{g}(\mathrm{x}_1, ..., x_n) \leqslant b$$
$$\mathrm{x}_i \geqslant 0$$

Tenemos el problema dual:

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

m vars
n restr.

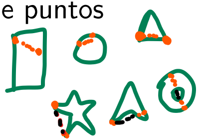
$$\text{s.a. } \mathrm{h}(\mathrm{y}_1, ..., y_m) \geqslant b$$
$$\mathrm{y}_j \leqslant 0$$

$$D(\bar{\lambda}) = \min_{\bar{x}} \mathcal{L}(x, \bar{\lambda})$$

- Si $\bar{x^*}$ es una solución factible del problema primal e $\bar{y^*}$ es una solución del dual entonces $f(\bar{x^*}) \geqslant D(\bar{y^*})$. $\min_x \max_y \geqslant \max_y \min_x$
- Si un problema no tiene un óptimo finito entonces el otro no es factible. $\nexists\ \bar{x}^*$ para $f \Rightarrow B_{DUAL} \neq \emptyset$

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 $z = \alpha x + (1-\alpha)y \in S$.



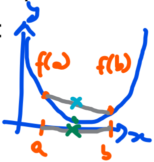
Obs: X_1, X_2 son dos conjuntos convexos, entonces,

- $X_1 \cap X_2$ 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, $f : M \to \mathbb{R}$. Entonces se dice que:

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



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

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, $f : M \to \mathbb{R}$ diferenciable, se dice que:
f es convexa en $M \Leftrightarrow \forall x, y \in M, f(y) \geqslant f(x) + \nabla f(x)(y - x)$
O bien, $(\nabla f(y) - \nabla f(x))(y - x) \geqslant 0$

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

Condiciones Suficientes de Optimalidad Global

$$\begin{gathered} opt\ \ f(x_1, ..., x_n) \\ \text{s.a. } \mathrm{g}_1(x_1, ..., x_n) = 0 \\ \vdots \\ \mathrm{g}_m(x_1, ..., x_n) = 0 \end{gathered}$$

con $m < n$ donde f y $g_i, i = 1, \ldots m$ 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.

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

$g_1(\bar{x}) = 0$

$\Updownarrow$

$\begin{cases} g_1(\bar{x}) \leq 0 \\ g_2(\bar{x}) \leq 0 \end{cases}$ con $g_2 = -g_1$

$g_3(\bar{x}) \geq 0$

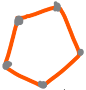
$\Updownarrow$

$g_4(\bar{x}) \leq 0$ con $g_4 = -g_3$

$$
\begin{aligned}
& opt\ c_1x_1 + \ldots + c_nx_n \\
\text{s.a. } & a_{11}x_1 + \ldots + a_{1n}x_n \leqslant b_1 \\
& \vdots \\
& a_{m1}x_1 + \ldots + a_{mn}x_n \leqslant b_m \\
& d_{11}x_1 + \ldots + d_{1n}x_n \geqslant e_1 \\
& \vdots \\
& d_{r1}x_1 + \ldots + d_{rn}x_n \geqslant e_r \\
& g_{11}x_1 + \ldots + g_{1n}x_n = b_1 \\
& \vdots \\
& g_{s1}x_1 + \ldots + g_{sn}x_n = b_s
\end{aligned}
$$

$\min\ c^T x$

s.t. $A \cdot x \leq b$



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.

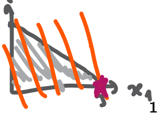
opt local = opt global

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

max $x_1^2 + x_2^2$
s.t $x_1 \geq x_2$

máx $x_1 + x_2$
s.t $|x_1| + |x_2| \leq 1$

máx $x_1 + x_2$
s.t $x_1 \geq 0$
$x_2 \geq 0$
$x_1 + 2x_2 \leq 3$

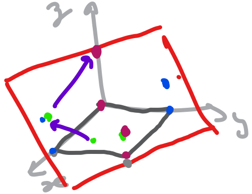


Iniciar: eligiendo una solución básica factible de una submatriz B de A, $|B| \neq 0$. *Iterar*:

1. Resolver el sistema $B\bar{x_B} = \bar{b}$. Entonces la solución será $\bar{x} = (\bar{x_B}, \bar{x_N})$.
2. 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 $z_k - c_k = mx\{z_j - c_j, j \in J\}$ donde J los índices asociados a las variables no básicas.
   - Si $z_k - c_k \leqslant 0$, 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...
3. Antes de construir una nueva solución conviene ver si existe $j \in J : z_k - c_k > 0, \bar{y}_j \leqslant 0$. 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:

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



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

Programación cuadrática

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

$$min_{x \in \mathbb{R}^d} \; \frac{1}{2} x^T Q x + c^T x$$

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

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:

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

$\min_{x} \mathcal{L}$

Y sustituyendo obtenemos la lagrangiana dual:

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