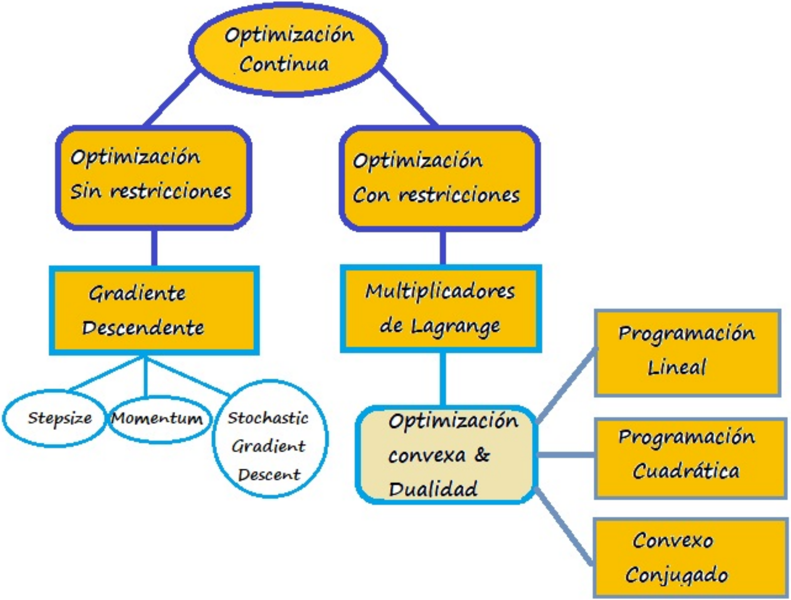
Clase 6/7: Optimización

Verónica Pastor, Martín Errázquin

Análisis Matemático para Inteligencia Artificial

14/04/2023

Ramas Principales de la Optimización



Optimización con restricciones

Dado el problema de optimización con restricciones:

función objetivo

sujeto a

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

campo escalar

mismo

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

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

comb. lineal de las m restricciones

multiplicadores de Lagrange

$$\mathscr{L}(\bar{\lambda}, \bar{x}) = f(\bar{x}) + \overbrace{\lambda_1 g_1(\bar{x}) + ... + \lambda_m g_m(\bar{x})}^{\text{comb. lineal de las m restricciones}}, \quad con\ \bar{\lambda} = \overbrace{(\lambda_1, ..., \lambda_m)}^{\text{multiplicadores de Lagrange}}$$

$$= f(\bar{x}) + (\lambda_1 \dots \lambda_m) \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix}$$

$Ax = b$

Teorema de condición necesaria de Lagrange

cant. de restricciones (m), sea menor que el Dom (f)

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

$$B = \{\bar{x} \in \underbrace{D}_{\subset \mathbb{R}^n} : g_i(\bar{x}) = 0, i = 1, m\}$$

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:

Pedir esto es equivalente a pedir que $\nabla g_1(\bar{x}^*) \ldots \nabla g_m(\bar{x}^*)$ son l.i

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

$$0 \neq [Jg] = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \ldots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1} & \ldots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

∇g_1, ∇g_m

s.s. las filas l.i

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 $\bar{x}^*$.

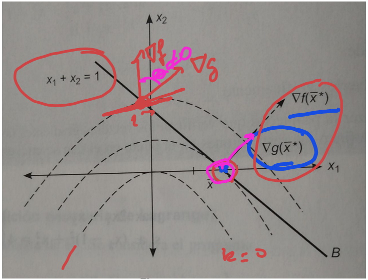
Ejemplo

función objetivo

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

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

$$g(x_1, x_2) = 0$$



Así definimos:

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

$$\nabla f(x^*) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right) = (2x_1, 2) = -\lambda \nabla g$$

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

$k=0 \quad x_2 = -\dfrac{x_1^2}{2}$

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

Como $\nabla f(x_0)$ es $\perp$ a la curva de nivel "k" en x_0

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 $\bar{x}^*$.

Optimización con restricciones de desigualdad

Condiciones necesarias de primer orden de Fritz-John

Dado el problema de optimización con restricciones:

$$\begin{aligned} opt\ f(x_1, ..., x_n) \\ \text{s.a. } \mathrm{g}_1(x_1, ..., x_n) \leqslant 0 \\ \vdots \\ \mathrm{g}_m(x_1, ..., x_n) \leqslant 0 \end{aligned} \tag{2}$$

donde $f : \mathbb{R}^n \to \mathbb{R}$ y $g_i : \mathbb{R}^n \to \mathbb{R}$ con $i = 1, \ldots m$. 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 $0 \leqslant \lambda_0, \lambda_i, i \in I$, y $g_i(\bar{x^*}) \leqslant 0, i = 1, m$.

Dualidad en Optimización: Problema dual

Dado el problema primal:

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

$$\text{s.a. } g(x_1, ..., x_n) \leqslant b$$
$$x_i \geqslant 0$$

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

iden las restricciones

$\lambda \in \mathbb{R}^m \quad m < n$

Tenemos el problema dual:

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

$$\text{s.a. } h(y_1, ..., y_m) \geqslant b$$
$$y_j \leqslant 0$$

$D: \tilde{Dom} \subset \mathbb{R}^m \longrightarrow \mathbb{R}$

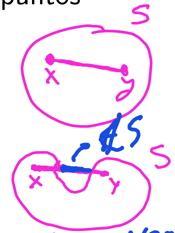
iden las restricciones

$\tilde{\lambda} \in \mathbb{R}^n \quad n < m$

- Si $\bar{x}^*$ es una solución factible del problema primal e $\bar{y}^*$ es una solución del dual entonces $f(\bar{x}^*) \geqslant D(\bar{y}^*)$.
- 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 $z = \alpha x + (1-\alpha)y \in S$.



∴ S no es conve

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 $M \Leftrightarrow \forall x, y \in M, f(y) \geqslant f(x) + \nabla f(x)(y-x)$ análoga a dif.
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 $y \in \mathbb{R}^n$.

H Matriz hessiana

$\mathcal{C}^2$ la función f, sus der. parciales de orden 1 y orden 2 existen y son cont.

Condiciones Suficientes de Optimalidad Global

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

dif. ó deriv. primeras continuas.

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.

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:

$$opt\ c_1x_1 + ... + c_nx_n$$

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

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. $A\bar{x} = \bar{b}; \bar{x} \geqslant 0$, donde $A \in \mathbb{R}^{m \times n}$, $m < n$, $rg(A) = m$.

Algoritmo del simplex

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:

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

forma cuadrática (*)

$$\text{s.a: } \overset{m\times d,\ d\times 1}{Ax} \leqslant \overset{m \times 1}{b} \quad \text{donde } A \in \mathbb{R}^{m\times d}, b \in \mathbb{R}^m, c \in \mathbb{R}^d.$$

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:

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

Derivando respecto a x, y despejando:

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

Recordar vimos que si una matriz es cuadrática ent. tiene inversa

Y sustituyendo obtenemos la lagrangiana dual:

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

Convexo Conjugado

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:

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

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.