

OPTIMIZACIÓN

Primer Cuatrimestre 2025

Práctica N° 3: Optimización con restricciones (Parte I).

Ejercicio 1 Considere el problema de minimizar la suma de las desviaciones absolutas,

$$\min_w \sum_{i=1}^m |y_i - x_i^T w|.$$

Muestre que se puede reformular como un problema de programación lineal.

Ejercicio 2 Considere el problema de regresión LASSO,

$$\min_w \frac{1}{2} \|y - Xw\|_2^2 + \lambda \|w\|_1,$$

Muestre que se puede reformular a un problema de tipo programación cuadrática con restricciones lineales. Sugerencia: considere las partes positiva y negativa de w .

Ejercicio 3 Considere el problema en el que se desea minimizar el costo del peor caso sobre un conjunto de funciones $f_i(x)$, es decir,

$$\min_{x \in X} \max_{i \in I} f_i(x).$$

Muestre que se puede reformular a un problema de optimización de tipo estándar con restricciones de desigualdad. Sugerencia: considere una variable que acote todas las funciones.

Ejercicio 4 Considere el problema de denoising por variación total, dado por

$$\min_x \frac{1}{2} \|y - x\|_2^2 + \lambda \sum_{i=1}^{n-1} |x_{i+1} - x_i|.$$

Muestre que se puede reformular a un problema de optimización cuadrática con restricciones lineales.

Ejercicio 5 Considere el problema de minimizar la suma de las pérdidas de Huber en una regresión robusta, es decir,

$$\min_w \sum_{i=1}^m \ell_\delta(r_i) \quad \text{con} \quad r_i = y_i - x_i^T w,$$

donde la función de pérdida de Huber se define como

$$\ell_\delta(r) = \begin{cases} \frac{1}{2} r^2, & \text{si } |r| \leq \delta, \\ \delta |r| - \frac{1}{2} \delta^2, & \text{si } |r| > \delta. \end{cases}$$

Muestre que se puede reformular a un problema convexo, con costo lineal y restricciones cuadráticas. Para ello, introduzca, para cada observación i , una variable auxiliar t_i que acote la pérdida $\ell_\delta(r_i)$, y además una variable s_i que represente $|r_i|$, con $r_i = y_i - x_i^T w$.

Ejercicio 6 Dado $v \in \mathbb{R}^n$ y el hiperplano $a^T x = b$ con $a \in \mathbb{R}^n$ y $b \in \mathbb{R}$, encuentre una solución analítica para el problema de la proyección ortogonal de v sobre el hiperplano.

$$\min_{x \in \mathbb{R}^n} \|x - v\|_2^2 \quad \text{sujeto a} \quad a^T x = b.$$

Ejercicio 7 Considere un sistema subdeterminado $Ax = b$ con $A \in \mathbb{R}^{m \times n}$ (donde $m < n$) y $b \in \mathbb{R}^m$ de modo que $Ax = b$ admite infinitas soluciones. Para seleccionar la solución de mínima norma, plantee el siguiente problema de optimización:

$$\min_{x \in \mathbb{R}^n} \|x\|_2^2 \\ \text{sujeto a} \quad Ax = b.$$

Muestre que este problema se reduce a la resolución de un sistema lineal

Ejercicio 8 Sea $X \in \mathbb{R}^{m \times n}$ la matriz de diseño y $y \in \mathbb{R}^m$ el vector de observaciones.

$$\min_{w \in \mathbb{R}^n} \|Xw - y\|_2^2 \quad \text{sujeto a} \quad \sum_{i=1}^n w_i = 1.$$

Muestre que el problema puede reducirse a la resolución de un sistema lineal.

Ejercicio 9 Dado el problema:

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s.a. : } \mathbf{Ax} = \mathbf{b}, \end{aligned} \tag{1}$$

con $\mathbf{A} \in \mathbb{R}^{m \times n}$, sea $\bar{\mathbf{x}}$ tal que $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$.

(a) Probar que (1) es equivalente a:

$$\min f(\bar{\mathbf{x}} + \mathbf{Bz}), \tag{2}$$

para cierta $\mathbf{B} \in \mathbb{R}^{n \times (n-m)}$. ¿Quién es \mathbf{B} ?

(b) Escribir las iteraciones de los métodos del gradiente y de Newton para (2) en función de las derivadas de f y de \mathbf{B} .

Ejercicio 10 Considerar el problema de programación cuadrática:

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{x}^t \mathbf{Q} \mathbf{x} - \mathbf{b}^t \mathbf{x} \\ \text{s.a. :} \quad & \mathbf{Ax} = \mathbf{c}, \end{aligned}$$

donde $\mathbf{Q} \in \mathbb{R}^{n \times n}$ y $\mathbf{A} \in \mathbb{R}^{p \times n}$ de rango $p < n$.

(a) Probar que si \mathbf{x}^* es un mínimo local, entonces es un mínimo global (sin asumir convexidad).

(b) Mostrar que el problema puede reducirse a la minimización sin restricciones de una función de $n - p$ variables.

(c) Despejar \mathbf{x}^* en términos de \mathbf{Q} , \mathbf{A} , \mathbf{b} y \mathbf{c} .

Ejercicio 11 Considere el problema de control discreto, donde para un x_0 dado tenemos una función de costo J , restricciones dadas por una dinámica f , y controles u_k que son las variables que podemos optimizar:

$$\begin{aligned} \min_{\{u_k\}_{k=0}^{N-1}} J(x_0, \{u_k\}) &= \sum_{k=0}^{N-1} L(x_k, u_k) + \phi(x_N) \\ \text{sujeto a} \quad x_{k+1} &= f(x_k, u_k), \quad k = 0, 1, \dots, N-1, \end{aligned}$$

Definimos el Hamiltoniano para cada instante k como:

$$H(x_k, u_k, \lambda_{k+1}) = L(x_k, u_k) + \lambda_{k+1}^T f(x_k, u_k).$$

Muestre que las condiciones de optimalidad conducen a las siguientes ecuaciones:

(a) Estacionalidad con respecto al control:

$$\frac{\partial H(x_k, u_k, \lambda_{k+1})}{\partial u_k} = 0.$$

(b) La ecuación adjunta:

$$\lambda_k = \frac{\partial H(x_k, u_k, \lambda_{k+1})}{\partial x_k}, \quad k = 0, \dots, N-1,$$

(c) La condición terminal

$$\lambda_N = \frac{\partial \phi(x_N)}{\partial x_N}.$$

Muestre que, para un x_0 dado, es posible integrar la dinámica hacia adelante en el tiempo para obtener $\{x_k\}$, resolver la ecuación adjunta hacia atrás en el tiempo para obtener $\{\lambda_k\}$, y de ese modo reducir el problema a un problema de optimización sin restricciones para las variables $\{u_k\}$.

Ejercicio 12 Considere el problema de entrenamiento de una red neuronal de L capas, formulado como un problema de optimización con restricciones de igualdad:

$$\begin{aligned} & \min_{\{a^{(l)}, W^{(l)}, b^{(l)}\}} \mathcal{L}(a^{(L)}, y) \\ & \text{sujeto a: } a^{(l)} - f^{(l)}\left(W^{(l)} a^{(l-1)} + b^{(l)}\right) = 0, \quad l = 1, \dots, L. \end{aligned}$$

donde:

- La activación de la capa 0 se define como $a^{(0)} = x$, donde x es la entrada.
- Para cada capa $l = 1, 2, \dots, L$, la activación se define mediante $a^{(l)} = f^{(l)}\left(W^{(l)} a^{(l-1)} + b^{(l)}\right)$, donde: $W^{(l)}$ y $b^{(l)}$ son los pesos y sesgos de la capa l y $f^{(l)}(\cdot)$ es la función de activación en la capa l .
- $\mathcal{L}(a^{(L)}, y)$ es la función de pérdida que mide el error entre la salida $a^{(L)}$ y el objetivo y .

Formule las condiciones de optimalidad dadas por el teorema de multiplicadores de Lagrange, y muestre que las distintas condiciones obtenidas:

- (a) Recuperan la dinámica de propagación hacia adelante de la red.
- (b) El error en la salida se “retropropaga” a través de las capas. Sugerencia: obtenga una relación recursiva entre una los multiplicadores $\lambda^{(l)}$ y $\lambda^{(l+1)}$.
- (c) Permiten calcular las derivadas del Lagrangiano con respecto a los pesos $W^{(l)}$ y los sesgos $b^{(l)}$ de manera explícita.

Nota: este método para calcular el gradiente se conoce como *backpropagation* y la presente formulación Lagrangiana se debe a Y. LeCun “A theoretical framework for backpropagation”, 1988.