
CONVEXIDAD Y OPTIMIZACIÓN

ENTREGA B

Christian Limbert Paredes Aguilera

Latex Source: <https://n9.c1/ua8c3>

CAPÍTULO 2

Ejercicio 1. Se considera el problemas

$$\begin{array}{ll} \text{minimiza} & x^2 + 1 \\ \text{sujeto a} & (x - 2)(x - 4) \leq 0, \end{array}$$

con $x \in \mathbb{R}$.

(a) Describe el conjunto accesible, el valor óptimo y las solución optima.

Solución:

(b) Escribe la función Lagrangiana $L(x, \lambda)$ y dibuja las funciones $x \mapsto L(x, \lambda)$ para algunos valores positivos de λ . Verifica que $p^* \geq \inf_x L(x, \lambda)$ para $\lambda \leq 0$. Obtén y dibuja la gráfica de la función dual de Lagrange g .

Solución:

(c) Describe el problema dual y comprueba que se trata de un problema de maximización cóncavo. Encuentra el valor dual óptimo d^+ y la solución dual óptima λ^* . ¿Se verifica la dualidad fuerte? Representa el conjunto $G \subset \mathbb{R}^2$.

Solución:

Ejercicio 2. ¿Cuáles son las condiciones de Slater asociadas al problema lineal

$$\begin{array}{ll} \text{minimiza} & x^T x \\ \text{sujeto a} & Ax = b? \end{array}$$

Se sabe que las condiciones de Slater implican que se verifica la dualidad fuerte. Describe esta situación para el problema lineal planteado. ¿Es posible aplicar las condiciones de Slater al problema 1?

Solución:

Ejercicio 3. Considera el problema de optimización

$$\begin{array}{ll} \text{minimiza} & e^{-x} \\ \text{sujeto a} & x^2/y \leq 0, \end{array}$$

con variables x e y , y dominio $D = \{(x, y) : y > 0\}$.

(a) **Comprueba que es un problema de optimización convexa. Encuentra el valor óptimo**

Solución: Para demostrar que el problema de optimización dado es convexo y encontrar el valor óptimo, necesitamos verificar dos cosas: la convexidad de la función objetivo y la convexidad de las restricciones.

Convexidad de la función objetivo.- La función objetivo es

$$f(x) = e^{-x}.$$

Para demostrar que es convexa, necesitamos mostrar que su segunda derivada es siempre no negativa ¹.

Empecemos calculando la primera derivada de $f(x)$ que se obtiene utilizando la regla de la cadena para la derivación. La derivada de e^u con respecto a u es e^u , y la derivada de $-x$ con respecto a x es -1 . Por lo tanto, la primera derivada de $f(x)$ es:

$$f'(x) = \frac{d}{dx}e^{-x} = -e^{-x}.$$

La segunda derivada de $f(x)$ se obtiene derivando $f'(x)$ con respecto a x . Nuevamente, utilizamos la regla de la cadena. La derivada de e^u con respecto a u es e^u , y la derivada de $-x$ con respecto a x es -1 . Por lo tanto, la segunda derivada de $f(x)$ es:

$$f''(x) = \frac{d}{dx}(-e^{-x}) = e^{-x}.$$

Sabiendo que una función es convexa si su segunda derivada es siempre no negativa. La segunda derivada de $f(x)$ es $f''(x) = e^{-x}$, que es siempre positiva para todo x en el dominio de los números reales, ya que la función exponencial e^{-x} es siempre positiva. Por lo tanto, la función objetivo $f(x) = e^{-x}$ es convexa.

Convexidad de las restricciones.- Sea la restricción,

$$g(x, y) = x^2/y \leq 0.$$

Ya que $y > 0$ (según $D = \{(x, y) : y > 0\}$), esta restricción implica que

$$x = 0.$$

Un conjunto que consiste en un solo punto (en este caso, el punto $x = 0$) es un conjunto convexo. Esto se debe a que para cualquier par de puntos en el conjunto (que en este caso son el mismo punto $x = 0$), el segmento de línea que los conecta también está en el conjunto. Por lo tanto, la restricción $g(x, y) = x^2/y \leq 0$ define un conjunto convexo.

Así, el problema de optimización dado es convexo.

Para encontrar el valor óptimo del problema de optimización dado, debemos evaluar la función objetivo en el punto que satisface la restricción. Es decir, debemos evaluar la función objetivo en $x = 0$. Haciendo esto, obtenemos:

$$f(0) = e^{-0} = e^0 = 1.$$

¹Una función $f : \mathbb{R} \rightarrow \mathbb{R}$ es convexa en un intervalo I si para todo $x \in I$, la segunda derivada $f''(x)$ existe y $f''(x) \geq 0$. En otras palabras, si la segunda derivada de una función es siempre no negativa en su dominio, entonces la función es convexa en ese dominio. Esto se debe a que la segunda derivada de una función mide la curvatura de la función, y una curvatura no negativa implica que la función es convexa.

Por lo tanto, el valor óptimo del problema de optimización dado es 1.

Este resultado se obtiene al considerar que la función objetivo $f(x)$ es decreciente² para $x \geq 0$ y la restricción implica que $x = 0$. Por lo tanto, el valor mínimo de $f(x)$ en el dominio dado es $f(0) = e^0 = 1$.

- (b) **Establece el problema dual y encuentra la solución óptima dual λ^* y el valor óptimo d^* . ¿Cuál es el salto de dualidad $p^* - d^*$? Representa el conjunto $G \subseteq \mathbb{R}^2$.**

Solución: El problema de optimización primal que nos has dado es el siguiente:

$$\begin{array}{ll} \text{minimiza} & e^{-x} \\ \text{sujeito a} & x^2/y \leq 0, \end{array}$$

con variables x e y , y dominio $D = \{(x, y) : y > 0\}$.

Para formular el problema dual, primero expresamos el problema primal en su forma estándar. La función objetivo es $f_0(x) = e^{-x}$ y la restricción es $f_1(x, y) = x^2/y \leq 0$. Entonces, el problema primal se puede escribir como:

$$\begin{array}{ll} \text{minimiza} & f_0(x) \\ \text{sujeito a} & f_1(x, y) \leq 0, \end{array}$$

El problema dual asociado es:

$$\begin{array}{ll} \text{maximiza} & -\inf_x (e^{-x} + \lambda x^2/y) \\ \text{sujeito a} & \lambda \geq 0, \end{array}$$

donde λ es la variable dual.

Dado que la restricción primal implica que $x = 0$, la función objetivo dual se simplifica a $-\inf_x e^{-x} = -1$. Por lo tanto, el problema dual se reduce a:

$$\begin{array}{ll} \text{maximiza} & -1 \\ \text{sujeito a} & \lambda \geq 0, \end{array}$$

La solución óptima dual es $\lambda^* = 0$ y el valor óptimo dual es $d^* = -1$.

El salto de dualidad es $p^* - d^* = 1 - (-1) = 2$, donde $p^* = 1$ es el valor óptimo primal.

El conjunto G es el conjunto de todos los (x, y) que satisfacen la restricción $x^2/y \leq 0$ con $y > 0$. Dado que la restricción implica que $x = 0$, el conjunto G es simplemente el eje y para $y > 0$, es decir, $G = \{(0, y) : y > 0\}$.

- (c) **¿Se cumple la condición de Slater para este problema?**

Solución: La condición de Slater es una condición de calificación de restricciones que garantiza la dualidad fuerte en problemas de optimización convexa. La condición de Slater requiere que exista al menos un punto estrictamente factible, es decir, un punto que satisfaga todas las restricciones de desigualdad de manera estricta.

En este problema de optimización, la restricción es $g(x, y) = x^2/y \leq 0$ con $y > 0$. Dado que $y > 0$, la restricción implica que $x = 0$. Por lo tanto, el conjunto factible es el conjunto de puntos $(0, y)$ para $y > 0$.

Dado que no hay ningún punto que satisfaga la restricción de manera estricta (es decir, con $x^2/y < 0$), la condición de Slater no se cumple para este problema de optimización.

²La función objetivo en este problema de optimización es $f(x) = e^{-x}$. Esta es una función exponencial decreciente para $x \geq 0$, lo que significa que a medida que x aumenta, $f(x)$ disminuye.

Ejercicio 4. Considera el problema:

$$\begin{array}{ll} \text{minimiza} & f_0(x) \\ \text{sujeto a} & f_i(x) \leq 0, \quad i = 1, \dots, m \end{array}$$

donde las funciones $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ son diferenciables y convexas. Muestra que la función

$$\phi(x) = f_0(x) + \alpha \sum_{i=1}^m \max\{0, f_i(x)\}^2,$$

es convexa (cuando $\alpha > 0$). Si \tilde{x} minimiza ϕ , muestra cómo obtener, a partir de \tilde{x} , un punto accesible para el problema dual. Encuentra la cota inferior correspondiente al problema primal.

Solución:

Ejercicio 5. Minimizando una función cuadrática. Considera el problema de minimizar una función cuadrática:

$$\text{minimiza } f(x) = (1/2)x^T P x + q^T x + r,$$

donde P es una matriz simétrica (no necesariamente en S_+^n).

- (a) Muestra que si $P \not\geq 0$ (es decir, la función no es convexa) entonces el problema no está acotado inferiormente.

Solución:

- (b) Si suponemos que $P \not\geq 0$, pero que la condición de optimalidad $Px^* = -q$ no tiene solución. Muestra que el problema vuelve a no estar acotado inferiormente.

Solución:

Ejercicio 6. Considera el problema siguiente:

$$\text{minimiza } f(x) = (1/2)(x_1^2 + \gamma x_2^2), \quad \gamma > 0.$$

- (a) Fija $(a, b) \in \mathbb{R}^2$. Calcula el valor de t que minimiza la función $f((a, b) - t\nabla f(a, b))$ en términos de a y b , claro.

Solución: Teniendo en cuenta que estamos trabajando en \mathbb{R}^2 y que $\gamma > 0$:

Consideremos la función

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}$$

definida por

$$f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2),$$

donde $x = (x_1, x_2)$ es un vector en \mathbb{R}^2 y $\gamma > 0$ es una constante.

Sabemos que el gradiente de una función es un vector que apunta en la dirección en la que la función aumenta más rápidamente. En otras palabras, si estamos en un punto en el espacio y queremos saber en qué dirección debemos movernos para aumentar el valor de la función lo más rápido posible, debemos movernos en la dirección del gradiente.

Para calcular el gradiente de esta función, tomamos las derivadas parciales de la función con respecto a cada una de las variables, x_1 y x_2 .

La derivada parcial de f con respecto a x_1 es la derivada de $\frac{1}{2}x_1^2$ con respecto a x_1 , manteniendo x_2 constante. Como la derivada de x_1^2 con respecto a x_1 es $2x_1$, la derivada parcial de f con respecto a x_1 es

$$x_1.$$

De manera similar, la derivada parcial de f con respecto a x_2 es la derivada de $\frac{1}{2}\gamma x_2^2$ con respecto a x_2 , manteniendo x_1 constante. Como la derivada de γx_2^2 con respecto a x_2 es $2\gamma x_2$, la derivada parcial de f con respecto a x_2 es

$$\gamma x_2.$$

Por lo tanto, el gradiente de f es

$$(x_1, \gamma x_2).$$

Así, el gradiente de la función en el punto (a, b) en \mathbb{R}^2 es

$$\nabla f(a, b) = (a, \gamma b).$$

Esto significa que si estás en el punto (a, b) en \mathbb{R}^2 , la dirección en la que la función $f(x)$ aumenta más rápidamente es en la dirección del vector $(a, \gamma b)$.

Consideremos ahora un paso de gradiente desde el punto (a, b) en la dirección opuesta al gradiente³. Este paso se puede representar como

$$(a, b) - t\nabla f(a, b) = (a, b) - t(a, \gamma b),$$

donde t es el valor que minimiza la función.

Vamos a calcular el valor de t que minimiza la función $f((a, b) - t\nabla f(a, b))$ ⁴. Para hacer esto, primero necesitamos expresar la función f en términos de t . Sustituyendo $(a, b) - t\nabla f(a, b)$ en f , obtenemos

$$\begin{aligned} f((a, b) - t\nabla f(a, b)) &= f((a, b) - t(a, \gamma b)) \\ &= f((a - ta, b - t\gamma b)) \\ &= \frac{1}{2}((a - ta)^2 + \gamma(b - t\gamma b)^2). \end{aligned}$$

Ahora, encontremos el valor de t que minimiza la función. Primero, tenemos la derivada de la función $f((a, b) - t\nabla f(a, b))$ con respecto a t :

$$\frac{df}{dt} = -(a^2 - 2a^2t + \gamma b^2 - 2\gamma^2 b^2 t).$$

Para encontrar el valor de t que minimiza la función, necesitamos igualar esta derivada a cero. Esto nos da la ecuación:

$$-(a^2 - 2a^2t + \gamma b^2 - 2\gamma^2 b^2 t) = 0.$$

³Recordemos que una forma de minimizar una función es a través de un proceso llamado "descenso de gradiente". La idea es comenzar en un punto inicial y luego moverse repetidamente en la dirección que disminuye la función más rápidamente, que es la dirección opuesta al gradiente. Entonces, cuando decimos que el punto (a, b) es "desplazado" por $-t\nabla f(a, b)$, lo que realmente queremos decir es que estamos moviendo el punto (a, b) una pequeña cantidad en la dirección que disminuye la función más rápidamente. El tamaño de este movimiento está determinado por t , y la dirección de este movimiento está determinada por el gradiente $\nabla f(a, b)$. Por lo tanto, $(a, b) - t\nabla f(a, b)$ representa el nuevo punto después de dar un paso de gradiente de tamaño t desde el punto (a, b) en la dirección opuesta al gradiente.

⁴El valor de t se calcula en la función f porque estamos buscando el valor que minimiza la función f después de un paso de gradiente.

Podemos simplificar esta ecuación distribuyendo el signo negativo a través de los términos en el paréntesis:

$$-a^2 + 2a^2t - \gamma b^2 + 2\gamma^2 b^2 t = 0.$$

Luego, reorganizamos los términos para agrupar los términos que contienen t :

$$2a^2t + 2\gamma^2 b^2 t = a^2 + \gamma b^2.$$

Factorizamos t del lado izquierdo de la ecuación:

$$t(2a^2 + 2\gamma^2 b^2) = a^2 + \gamma b^2.$$

por lo tanto, para cualquier par de puntos $(a, b) \in \mathbb{R}^2$ y $\gamma > 0$ el valor de t que minimiza la función $f((a, b) - t\nabla f(a, b))$ es

$$t = \frac{a^2 + \gamma b^2}{2a^2 + 2\gamma^2 b^2}.$$

(b) Crea un programa en MATLAB o Python tal que:

- Tenga como entrada: el parámetro γ , una tolerancia ϵ , un número máximo de iteraciones N y un punto inicial $(x_1^{(0)}, x_2^{(0)})$.
- Aplique el método del gradiente al problema, utilizando paso exacto (obtenido en el apartado (a)).
- Que pare cuando llegue al número máximo de iteraciones o cuando la precisión sea menor que ϵ .
- El problema debe devolver la última aproximación y el número de iteraciones utilizado, así como un mensaje que informe de los posibles fallos producidos y de la condición de parada utilizada.

(c) Si te sientes con ganas, añade la funcionalidad de que al terminar represente el conjunto inicial S , las curvas de nivel por las que pasa cada aproximación, destacando el punto aproximado con un dot.

Solución de (b) y (c):

```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.colors import LogNorm

def gradient_descent(gamma, epsilon, N, initial_point):
    # Item 1: La funcion toma como entrada el parametro gamma,
    # una tolerancia epsilon, un numero maximo de iteraciones N
    # y un punto inicial (x1(0), x2(0)).

    # Definimos la funcion y su gradiente
    f = lambda x: 0.5 * (x[0]**2 + gamma*x[1]**2)
    grad_f = lambda x: np.array([x[0], gamma*x[1]])
```

```

# Inicializamos el punto y el numero de iteraciones
x = np.array(initial_point)
num_iterations = 0

# Creamos una lista para almacenar todas las aproximaciones
approximations = [x]

# Bucle principal del metodo del gradiente
while num_iterations < N:
    # Item 2: Aplicamos el metodo del gradiente al problema,
    # utilizando paso exacto.
    # Calculamos el gradiente
    gradient = grad_f(x)

    # Calculamos el tamano del paso
    t = (x[0]**2+gamma*x[1]**2)/(2*x[0]**2+2*gamma**2*x[1]**2)

    # Actualizamos el punto
    x = x - t * gradient

    # Anadimos la nueva aproximacion a la lista
    approximations.append(x)

    # Item 3: El bucle continua hasta que se alcanza el numero
    # maximo de iteraciones o hasta que la norma del gradiente
    # es menor que epsilon.
    if np.linalg.norm(gradient) < epsilon:
        print("El_algoritmo_convergio_con_exito.")
        break

    # Actualizamos el numero de iteraciones
    num_iterations += 1

# Item 4: La funcion devuelve la ultima aproximacion al minimo
# de la funcion y el numero de iteraciones que se utilizaron.
# Tambien se imprime un mensaje que informa de los posibles
# fallos producidos y de la condicion de parada que se utilizo.
if num_iterations == N:
    print("El_algoritmo_convergio_en_el_numero_maximo_de_iter.")
else:
    print("El_algoritmo_convergio_con_exito.")

# Creamos una cuadrícula de puntos en el espacio 2D
y = np.linspace(-10, 10, 400)
x = np.linspace(-10, 10, 400)
X, Y = np.meshgrid(x, y)

# Calculamos los valores de la funcion en cada punto de la
# cuadrícula
Z = 0.5 * (X**2 + gamma*Y**2)

# Creamos una figura y un conjunto de ejes en matplotlib
fig, ax = plt.subplots()

```

```
# Dibujamos las curvas de nivel de la funcion
ax.contour(X, Y, Z,
           levels=np.logspace(0, 5, 35),
           norm=LogNorm(),
           cmap=plt.cm.jet)

# Convertimos la lista de aproximaciones en un array de
# numpy para facilitar su manejo
approximations = np.array(approximations)

# Dibujamos los puntos aproximados
ax.plot(approximations[:, 0], approximations[:, 1], 'ko')

# Mostramos la figura
plt.show()

return x, num_iterations
```