



# Optimización Matemática

Grado en Matemática Aplicada

PhD. Thomas Ian Ashley  
Departamento de Métodos Cuantitativos

Optimización Matemática © 2025 by Thomas Ian Ashley is licensed under CC BY-NC-SA 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-sa/4.0/>



Universidad  
**LOYOLA**





## Índice general

### I Tema 1 - Fundamentos de la Optimización Matemática

<b>1</b>	<b>Fundamentos de la Optimización Matemática</b> .....	<b>7</b>
1.1	Introducción	7
1.2	Convexidad	11
1.3	Separación	18
1.3.1	Variantes de Farkas .....	26
1.4	Problemas	28

### II Tema 2 - Optimización sin restricciones

<b>2</b>	<b>Optimización sin restricciones</b> .....	<b>33</b>
2.1	Introducción	33
2.2	Condiciones de Optimalidad	33
2.3	Métodos Iterados	37
2.3.1	Método de Newton .....	37
2.3.2	Método de gradiente .....	38
2.4	Problemas	43

## III

## Tema 3 - Optimización restringida

<b>3</b>	<b>Optimización restringida</b> .....	<b>47</b>
<b>3.1</b>	<b>Introducción</b>	<b>47</b>
<b>3.2</b>	<b>Restricciones de igualdad</b>	<b>48</b>
3.2.1	Lagrange .....	49
3.2.2	Lagrange .....	50
3.2.3	Condiciones necesarias/suficientes .....	51
3.2.4	Sensibilidad .....	54
<b>3.3</b>	<b>Restricciones de desigualdad</b>	<b>55</b>
3.3.1	Karush-Kuhn-Tucker .....	57
3.3.2	Condiciones Suficientes .....	60
<b>3.4</b>	<b>Problemas</b>	<b>63</b>

## IV

## Tema 4 - Dualidad Lagrangiana

<b>4</b>	<b>Dualidad</b> .....	<b>67</b>
<b>4.1</b>	<b>Introducción</b>	<b>67</b>
4.1.1	No solución .....	73
4.1.2	Explicación Geométrica .....	74
<b>4.2</b>	<b>Problemas</b>	<b>77</b>

## V

## Práctica 1

<b>5</b>	<b>Práctica 1</b> .....	<b>81</b>
<b>5.1</b>	<b>Introducción</b>	<b>81</b>
<b>5.2</b>	<b>Código</b>	<b>81</b>
<b>5.3</b>	<b>Código</b>	<b>81</b>
<b>5.4</b>	<b>Preguntas</b>	<b>82</b>



# Tema 1 - Fundamentos de la Optimización Matemática

part. 1

## **1** Fundamentos de la Optimización Matemática ..... 7

- 1.1 Introducción
- 1.2 Convexidad
- 1.3 Separación
- 1.4 Problemas

chapter.2section.2.1section.2.2section.2.3subsection.2.3.1subsec





# 1. Fundamentos de la Optimización Matemática

## 1.1 Introducción

La optimización matemática es el proceso de la *formulación* y *resolución* de un problema matemático de la forma:

$$\begin{aligned} &\text{maximizar} && f(x) \\ &\text{sujeto a} && g_i(x) \leq b_i, \quad i = 1, 2, \dots, m, \\ & && h_j(x) = c_j, \quad j = 1, 2, \dots, k \\ & && x \geq 0. \end{aligned}$$

donde  $f(x)$ ,  $g_i(x)$  y  $h_j$  son funciones con  $n$  variables  $x \in \mathbb{R}^n$ .

**Nota** Resolver  $\max f(x)$  es igual a resolver  $\min -f(x)$ .

Aquí, la función  $f(x)$  se llama la **función objetivo** y  $x = (x_1, \dots, x_n)^T$  las **variables de decisión**. Las funciones  $g_i(x)$  y  $h_j(x)$  son las **restricciones** (también **ligaduras**) del problema, donde  $g_i(x)$  son restricciones de desigualdad, y  $h_j(x)$  son restricciones de igualdad.

■ **Ejemplo 1.1** Una empresa planea gastar 10000 euros en publicidad. Se sabe que un minuto de publicidad en televisión cuesta 3000 euros y 1000 euros en la radio. Si la empresa compra  $x$  minutos de publicidad en televisión e  $y$  minutos en la radio, su ingreso, en euros, está dado por  $-2x^2 - y^2 + xy + 8x + 3y$ . ¿Cómo puede la empresa maximizar sus ingresos?

Las variables de decisión del problema son:

$x$ : minutos que compra la empresa en televisión

$y$ : minutos que compra la empresa en radio

El objetivo es maximizar los ingresos,  $Z(x, y) = -2x^2 - y^2 + xy + 8x + 3y$ , y la restricción es gastar en total 10000 en publicidad  $3000x + 1000y = 10000$ . Por tanto, tenemos el problema:

$$\begin{aligned} \text{maximizar} \quad & Z(x, y) = -2x^2 - y^2 + xy + 8x + 3y \\ \text{sujeto a} \quad & 3000x + 1000y = 10000 \\ & x, y \geq 0. \end{aligned}$$

Diremos que  $x = (x_1, \dots, x_n)^T$  es una solución **factible** si cumple todas las restricciones  $g_i(x)$  y  $h_j(x)$ , y definimos el conjunto de todas las soluciones factibles, la **región factible**, por  $\Omega$ .

En problemas sencillos, graficar el conjunto de restricciones puede ayudar en su resolución.

■ **Ejemplo 1.2** Consideremos el siguiente problema de optimización

$$\begin{aligned} \text{maximizar} \quad & f(x) = x_1 \\ \text{sujeto a} \quad & x_1^2 + x_2^2 \leq 4, \\ & x_1 \geq 1. \end{aligned}$$

Podemos graficar sus restricciones en 2D e identificar la región factible del problema:

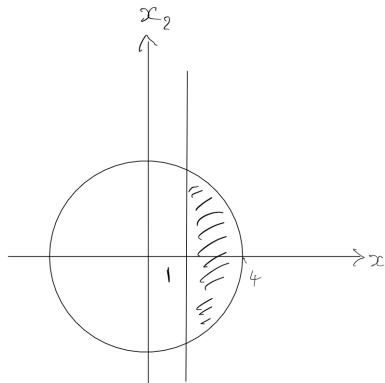


Figura 1.1: Región factible

En este ejemplo sencillo, la función objetivo es maximizar  $x_1$ , y desde la gráfica de la región factible, podemos identificar que el valor óptimo es  $x^* = (2, 0)$ .

■ **Ejemplo 1.3** En este ejemplo, consideramos un problema de optimización más complicado.

$$\begin{aligned} \text{maximizar} \quad & 3x_1 + 5x_2 = Z \\ \text{sujeto a} \quad & x_1 \leq 4, \\ & 3x_1 + 2x_2 \leq 18, \\ & 2x_2 \leq 12 \\ & x_1, x_2 \geq 0. \end{aligned}$$

Dibujamos la región factible:

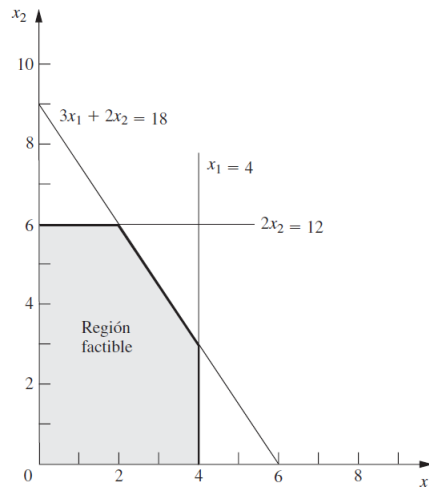


Figura 1.2: Región Factible

La función objetivo en este caso es un plano, y para ver su valor en la región factible, podemos proyectar su valor en líneas de nivel.

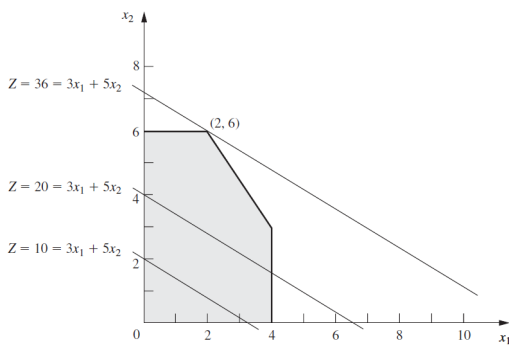


Figura 1.3: Valor óptimo

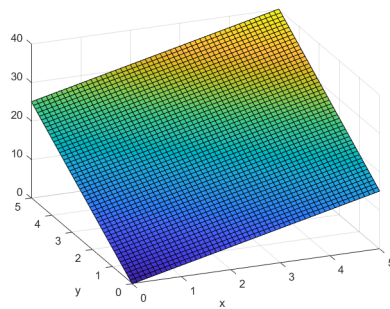


Figura 1.4: Función objetivo

El vector óptimo del problema se denota por  $x^*$ , y el valor óptimo correspondiente de la función objetivo por  $f(x^*)$ . Si el problema no tiene restricciones, se llama un problema de optimización *sin restricciones*.

El resultado principal utilizado para conocer si un problema de optimización tiene solución es el Teorema de Weierstrass.

**Teorema 1.1.1 — de Weierstrass.**

Sea  $f(x)$  una función continua definida sobre un conjunto compacto (cerrado y acotado)  $K \subseteq \mathbb{R}^n$ . Entonces el problema de optimización

$$\begin{cases} \text{Optimizar } f(x) \\ \text{Sujeto a } x \in K \end{cases}$$

tiene al menos una solución para ambos objetivos de minimización y maximización, es decir

$$\exists x_{\min}^*, x_{\max}^* \in K : \begin{cases} f(x_{\min}^*) = \min_{x \in K} f(x) \\ f(x_{\max}^*) = \max_{x \in K} f(x) \end{cases}$$

Un problema de optimización puede no tener solución, tener única solución, o tener varias. Por ejemplo con las funciones  $\sin(x)$  o  $x^2$ . También podemos diferenciar entre óptimos **locales** y **globales**.

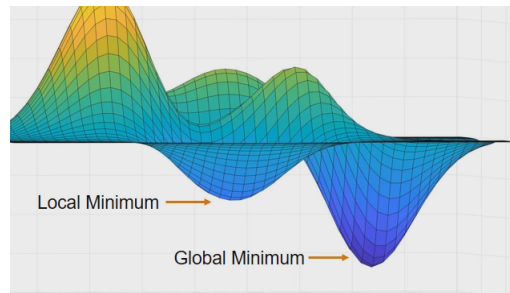


Figura 1.5: Óptimos locales y globales (fuente: Mathworks.com)

La optimización matemática, también conocida como la Investigación Operativa, engloba el análisis y resolución de problemas en el mundo real, con métodos y técnicas variados.

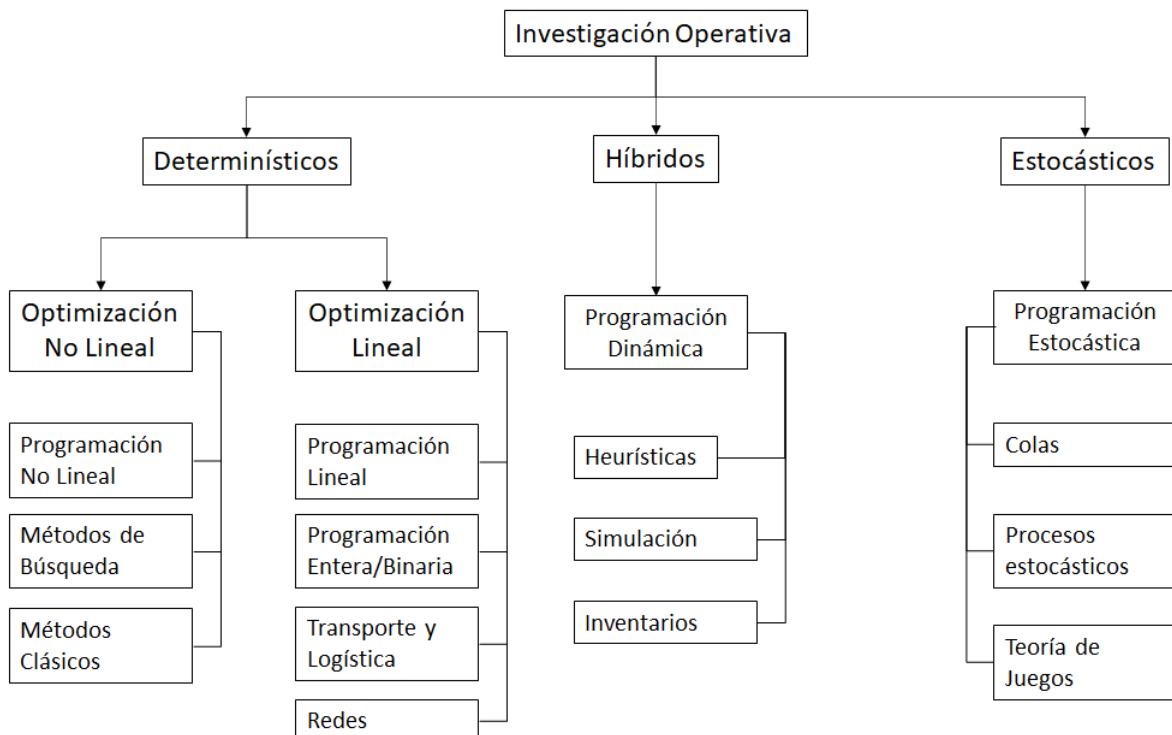


Figura 1.6: Investigación Operativa

La asignatura *Optimización Matemática* se concentra en definir conceptos fundamentales de la investigación operativa, y la definición y resolución de problemas de optimización *no lineales*. La teoría y resolución de problemas de optimización *lineales* será considerado en la asignatura

## 1.2 Convexidad

Una consideración especial en la resolución de problemas de optimización es la *convexidad* de la función objetivo y el sistema de restricciones. En este capítulo, consideremos las definiciones y teoremas importantes sobre la convexidad de conjuntos y funciones, y como pueden afectar la resolución de un problema de optimización.

### Definición 1.2.1 — Convexidad - Conjuntos.

Un conjunto  $C \subset \mathbb{R}^n$  es convexo si para todo  $a, b \in C$  tenemos que  $(1 - \lambda)a + \lambda b \in C, \forall \lambda \in [0, 1]$ .

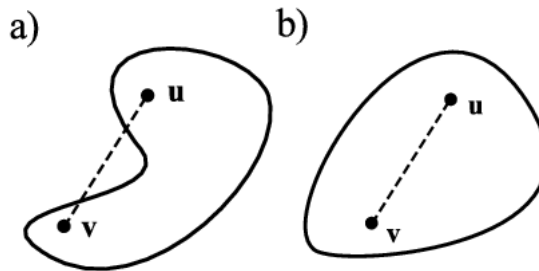


Figura 1.7: Convexidad de conjuntos

**Nota:** El conjunto vacío  $\emptyset$  es convexo.

### ■ Ejemplo 1.4

Demuestra que  $C = \{u \in \mathbb{R}^2 | u \cdot v \leq 9\}$  es convexo.

Si  $C$  es convexo, tenemos que por cualquier  $u, w \in C, t \in [0, 1], tu + (1 - t)w \in C$ . Entonces sean  $u, w \in C, t \in [0, 1]$ , queremos demostrar que  $(tu + (1 - t)w) \cdot v \in C$ .

$$\begin{aligned} (tu + (1 - t)w) \cdot v &= tu \cdot v + (1 - t)w \cdot v \\ &\leq t(9) + (1 - t)(9) \\ &\leq 9 \end{aligned}$$

Por lo tanto,  $(tu + (1 - t)w) \cdot v \in C \forall u, v$  y  $C$  es convexo. ■

### ■ Ejemplo 1.5

Sea  $A \subset \mathbb{R}^n$  y  $c \in \mathbb{R}$ , donde  $A$  es convexo. ¿Es  $cA = \{z \in \mathbb{R}^n | z = cx, x \in A\}$  convexo?

Sean  $z_1, z_2 \in cA$ . Entonces  $z_1 = cx_1, z_2 = cx_2 \in cA$ , donde  $x_1, x_2 \in A$ . Si  $\lambda \in [0, 1]$ , por la definición de convexidad tenemos:

$$\begin{aligned} \lambda z_1 + (1 - \lambda)z_2 &= \lambda cx_1 + (1 - \lambda)cx_2 \\ &= c(\lambda x_1 + (1 - \lambda)x_2) \in cA \Rightarrow \text{convexo.} \end{aligned}$$

ya que  $\lambda x_1 + (1 - \lambda)x_2 \in A$ , por ser  $A$  convexo. ■

**Lema 1.2.1**

Sean  $S_1, S_2 \subseteq \mathbb{R}^n$  convexos, entonces

- $S_1 \cap S_2$  es convexo,
- $S_1 + S_2 = \{x_1 + x_2 : x_1 \in S_1, x_2 \in S_2\}$  es convexo,
- $S_1 \ominus S_2 = \{x_1 - x_2 : x_1 \in S_1, x_2 \in S_2\}$  es convexo.

Demostramos la propiedad para la intersección de conjuntos convexos:

**Demostración 1.1**

Sea  $A_i \subset \mathbb{R}^n$   $i \in I$  una familia de conjuntos convexos. Si  $\bigcap_{i \in I} A_i = \emptyset$ , ya estaría demostrado.

Si  $\bigcap_{i \in I} A_i \neq \emptyset$ , sean  $x_1, x_2 \in \bigcap_{i \in I} A_i$ . Tenemos  $x_1, x_2 \in A_i, \forall i \in I$ , y por ser los conjuntos  $A_i$  convexos,

$$\lambda x_1 + (1 - \lambda)x_2 \in A_i \quad \forall i \in I, \forall \lambda \in (0, 1),$$

con lo que

$$\lambda x_1 + (1 - \lambda)x_2 \in \bigcap_{i \in I} A_i \quad \forall \lambda \in (0, 1),$$

esto es,  $\bigcap_{i \in I} A_i$  es un conjunto convexo.

Esta propiedad está representada en Figura 1.8.

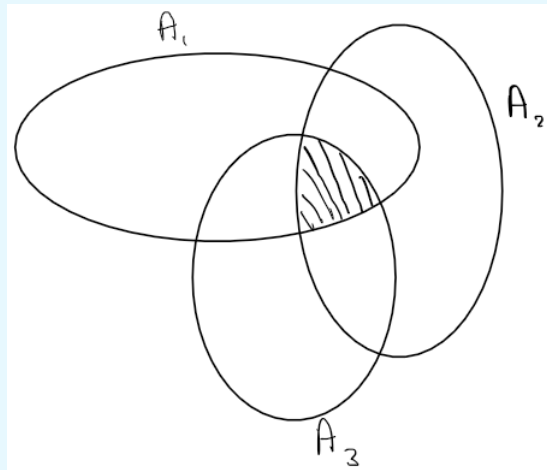


Figura 1.8: Combinación Convexa

**■ Ejemplo 1.6**

Si  $A$  y  $B$  son dos subconjuntos convexos no vacíos de  $\mathbb{R}^n$ , se define su suma como el conjunto

$$A + B = \{z \in \mathbb{R}^n \mid z = x + y, x \in A, y \in B\}.$$

Demuestra que  $A + B$  es un conjunto convexo.

Sean  $z_1, z_2 \in A + B$ . Se tiene que  $z_1 = x_1 + y_1$  y  $z_2 = x_2 + y_2$ , con  $x_1, x_2 \in A, y_1, y_2 \in B$ . Si  $t \in [0, 1]$ ,

$$\begin{aligned} tz_1 + (1 - t)z_2 &= t(x_1 + y_1) + (1 - t)(x_2 + y_2) \\ &= (tx_1 + (1 - t)x_2) + (ty_1 + (1 - t)y_2) \in A + B \end{aligned}$$

ya que  $tx_1 + (1 - t)x_2 \in A$  por ser convexo, y  $ty_1 + (1 - t)y_2 \in B$  por ser convexo. ■

**Definición 1.2.2 — Convexidad - Funciones.**

Diremos que la función  $f$  es convexa en  $C$  convexo si para todo  $x_1, x_2 \in C$ ,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad \forall \lambda \in [0, 1]$$

Si la desigualdad es estricta para  $x_1 \neq x_2, \lambda \in (0, 1)$  diremos que  $f$  es estrictamente convexa. Diremos que  $f$  es (estrictamente) cóncava si  $f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2) \quad \forall \lambda \in [0, 1]$ .

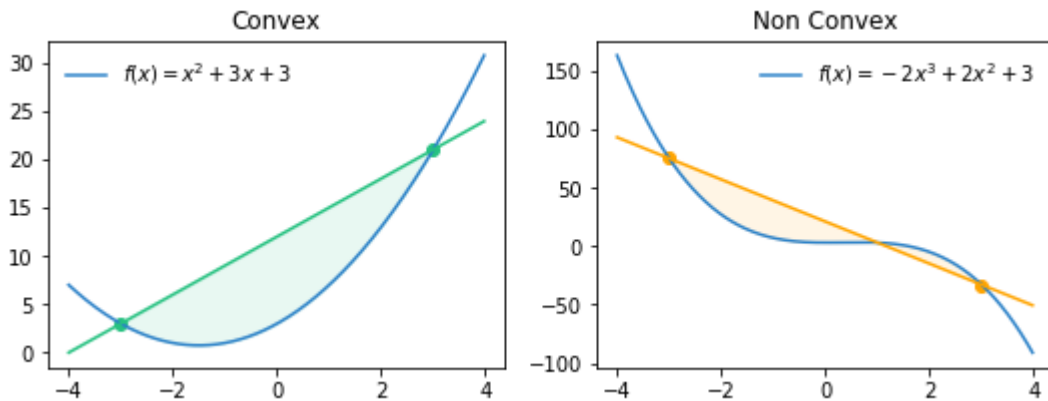


Figura 1.9: Convexidad de funciones

**■ Ejemplo 1.7**

Pruébese que la función  $f: \mathbb{R}^n \rightarrow \mathbb{R}, f(x) = a^T x + b, a, x \in \mathbb{R}^n, b \in \mathbb{R}$  es cóncava y convexa a la vez.

Sean  $x_1, x_2 \in \mathbb{R}^n$  y  $\lambda \in [0, 1]$ . Entonces, tenemos  $f(x_1) = a^T x_1 + b, f(x_2) = a^T x_2 + b$ . Si  $f$  es convexa y cóncava a la vez,  $f(\lambda x_1 + (1 - \lambda)x_2) = \lambda f(x_1) + (1 - \lambda)f(x_2)$ . Tenemos:

$$\begin{aligned} f(\lambda x_1 + (1 - \lambda)x_2) &= a^T (\lambda x_1 + (1 - \lambda)x_2) + b \\ &= \lambda a^T x_1 + (1 - \lambda)a^T x_2 + (1 + \lambda - \lambda)b \\ &= \lambda(a^T x_1 + b) + (1 - \lambda)(a^T x_2 + b). \end{aligned}$$

$\Rightarrow f$  es convexa y cóncava. ■

Una forma para comprobar la convexidad (o concavidad) de una función  $f(x) \quad x \in \mathbb{R}^n$  es calcular su Hessiana.

**Convexa** si la hessiana de  $f$  es semidefinida positiva

**Cóncava** si la hessiana de  $f$  es semidefinida negativa

**■ Ejemplo 1.8**

La función  $f(x) = x^2 + 2$  es una función de una variable. Usando la prueba de la segunda derivada, podemos identificar si es convexa o cóncava.  $f'' = 2 > 0$ , lo cual indica que la función es convexa. ■

**■ Ejemplo 1.9**

Consideremos la función  $f(x, y) = 2x^2 + 3x + 3y^2 + 1$ . ¿Es la función convexa en  $(x, y) \in \mathbb{R}^2$ ?

La hessiana de la función es  $Hf(x, y) = \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix}$ . Para determinar si la matriz es semidefinida

positiva o negativa, calculamos sus valores propios.  $\det(Hf(x,y) - \lambda I) = (4 - \lambda)(6 - \lambda)$ , igualando a cero  $\Rightarrow \lambda = 4$  o  $\lambda = 6$ . Por tener ambos valores propios positivos, la matriz es semidefinida positiva, y por lo tanto la función  $f(x,y)$  es convexa.

**Nota** Una matriz es semidefinida positiva cuando sus valores propios  $\lambda_i \geq 0 \forall i$ , semidefinida negativa cuando  $\lambda_i \leq 0 \forall i$  e indefinida cuando  $\exists \lambda_i < 0$  y  $\lambda_i > 0$ .

Otra forma de comprobar que una función es convexa es por su determinante  $D = f_{xx}f_{yy} - f_{xy}^2$ :

- Si  $D > 0$  y  $f_{xx}(x,y) > 0$  es convexa
- Si  $D > 0$  y  $f_{xx}(x,y) < 0$  es cóncava
- Si  $D < 0$  la matriz es indefinida

**Definición 1.2.3**

Sean los vectores  $a_1, \dots, a_k, b \in \mathbb{R}^n$ , diremos que  $b$  es

**Combinación lineal** de  $a_1, \dots, a_k$  si existen  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$  tal que  $b = \sum_{j=1}^k \lambda_j a_j$ ,

**Combinación afín** de  $a_1, \dots, a_k$  si existen  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$  con  $\sum_{j=1}^k \lambda_j = 1$  tal que  $b = \sum_{j=1}^k \lambda_j a_j$ ,

**Combinación convexa (estricta)** de  $a_1, \dots, a_k$  si existen  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$  con  $\sum_{j=1}^k \lambda_j = 1$  y  $\lambda_j \geq 0 (\lambda_j > 0)$  tal que  $b = \sum_{j=1}^k \lambda_j a_j$ .

■ **Ejemplo 1.10**

Consideramos los vectores  $B = \langle 1, 1 \rangle$  y  $C = \langle 2, 1 \rangle$  en 1.10.

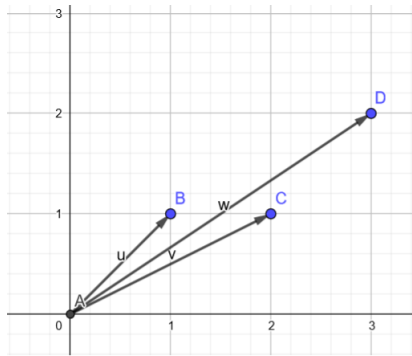


Figura 1.10: Combinación Lineal/afín

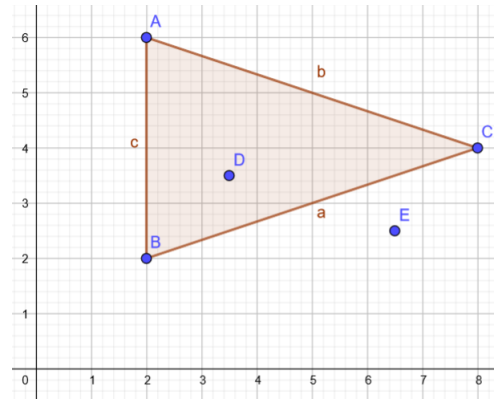


Figura 1.11: Combinación Convexa

Una combinación lineal de los vectores es  $1B + 1C = D = \langle 3, 2 \rangle$ . Una combinación afín puede ser  $\frac{1}{2}B + \frac{1}{2}C = \langle 1.5, 1 \rangle$ .

Ahora consideramos tres vectores  $A = \langle 2, 6 \rangle, B = \langle 2, 2 \rangle$  y  $C = \langle 8, 4 \rangle$  en 1.11. Vemos que la combinación afín  $E = \frac{-1}{4}A - \frac{3}{4}B + \frac{1}{2}C$  nos da un vector fuera de la región entre los tres puntos  $A, B$  y  $C$ . La combinación convexa  $D = \frac{1}{4}A + \frac{1}{2}B + \frac{1}{4}C$  nos da un vector dentro de la región.

**Definición 1.2.4**

Sean  $S \subseteq \mathbb{R}^n$  y  $\mathcal{C}_S$  la familia de los conjuntos convexos de  $\mathbb{R}^n$  que contienen a  $S$ . Llamaremos **envolvente convexa** de  $S$ ,  $C_0(S)$ , a la intersección de todos los conjuntos de  $\mathcal{C}_S$ . Es decir

$$C_0(S) = \left\{ x : x = \sum_{i=1}^k \lambda_i x_i, x_i \in S, \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1 \right\} \text{ con } k = 1, 2, \dots$$

Es decir, la envolvente convexa de  $S$  es el conjunto de puntos que pueden escribirse como combinación lineal convexa de una cantidad finita de puntos de  $S$ .

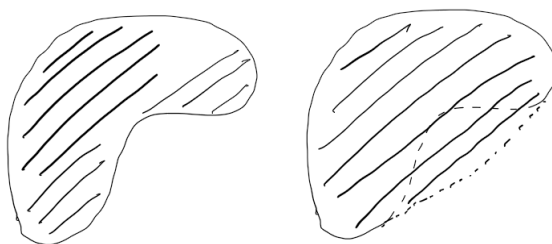


Figura 1.12: Envolvente Convexa

### demostración quitada

**Nota**  $A \subset C_0(A)$ .

**Nota** Si  $A \subset B$ , entonces  $C_0(A) \subset C_0(B)$ .

**Nota**  $C_0(A)$  es convexo.

### Demostración 1.2

Demostramos que  $C_0(A)$  es convexo.

Sean  $x, y \in C_0(S)$ , tenemos:

$$x = \sum_{i=1}^k \lambda_i x_i \quad y = \sum_{i=1}^l \beta_i y_i$$

con

$$\begin{aligned} 0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^k \lambda_i = 1, \quad x_i \in S, \quad i = 1, \dots, k, \\ 0 \leq \beta_j \leq 1, \quad \sum_{j=1}^l \beta_j = 1, \quad y_j \in S, \quad j = 1, \dots, l, \end{aligned}$$

Si  $0 \leq t \leq 1$

$$tx + (1-t)y = \sum_{i=1}^k t\lambda_i x_i + \sum_{j=1}^l (1-t)\beta_j y_j$$

ya que  $0 \leq t\lambda_i \leq 1, i = 1, \dots, k$  y  $0 \leq (1-t)\beta_j \leq 1, j = 1, \dots, l$ , tenemos

$$\begin{aligned} \sum_{i=1}^k t\lambda_i + \sum_{j=1}^l (1-t)\beta_j &= t \sum_{i=1}^k \lambda_i + (1-t) \sum_{j=1}^l \beta_j \\ &= t(1) + (1-t)(1) \\ &= 1 \end{aligned}$$

$\Rightarrow tx + (1-t)y \in C_0(S)$ . Entonces es convexo. ■

**Proposición 1.2.1**

Sean  $A, B \subset \mathbb{R}^n$  no vacíos.

1. Si  $A$  es convexo, entonces  $A = C_0(A)$ ,
2.  $C_0(A) + C_0(B) = C_0(A + B)$ .

**Demostración 1.3**

1)

Ya sabemos que  $A \subset C_0(A)$ . Recíprocamente, supongamos que  $A$  es convexo y veamos que  $C_0(A) \subset A$ . Si  $x \in C_0(A)$ , entonces existen  $x_1, \dots, x_k \in A$ ,  $\lambda_1, \dots, \lambda_k \geq 0$  con  $\sum_{i=1}^k \lambda_i = 1$  y  $\sum_{i=1}^k \lambda_i x_i = x$ . Es decir, que  $x$  es una combinación lineal convexa de  $x_i$ . Razonamos por inducción sobre  $k$ :

- Si  $k = 1$ , es evidente que  $x \in A$ .
- Si  $k = 2$ , entonces  $x \in [x_1, x_2] \subset A$  (por ser  $A$  convexo).
- Supongamos por hipótesis de inducción que si  $y \in C_0(A)$  se escribe como combinación lineal de *menos* de  $k$  puntos de  $A$ . Tomemos  $x \in C_0(A)$ , y consideremos valores de  $\lambda_k \in [0, 1]$ : Si  $\lambda_k = 0$ , entonces aplicamos a  $x$  nuestra hipótesis de inducción, y  $x \in A$  (por ser combinación lineal convexa de  $k - 1$  puntos). Si  $\lambda_k = 1$ , entonces  $\lambda_1 = \dots = \lambda_{k-1} = 0$  luego  $x$  es combinación lineal convexa de un punto de  $A$ , entonces  $x \in A$ . Si  $\lambda_k \in (0, 1)$ , entonces,

$$x = \sum_{i=1}^k \lambda_i x_i = (1 - \lambda_k) \sum_{i=1}^{k-1} \frac{\lambda_i}{1 - \lambda_k} x_i + \lambda_k x_k.$$

donde vemos que resulta una combinación lineal convexa de  $x_k$  y  $\sum_{i=1}^{k-1} \frac{\lambda_i}{1 - \lambda_k} x_i$ . Como  $A$  es convexo, solo necesitamos probar que  $\sum_{i=1}^{k-1} \frac{\lambda_i}{1 - \lambda_k} x_i \in A$  para que  $x$  esté en  $A$ . Pero  $\sum_{i=1}^{k-1} \frac{\lambda_i}{1 - \lambda_k} x_i$  es combinación lineal convexa de  $x_1, \dots, x_{k-1} \in A$ , porque  $\sum \frac{\lambda_i}{1 - \lambda_k} = 1$ . E.g.: si  $\lambda_k = 0.5 \Rightarrow \sum \lambda_i = 0.5 \Rightarrow \frac{0.5}{1 - 0.5} = 1$

Por el hipótesis de inducción,  $\sum_{i=1}^{k-1} \frac{\lambda_i}{1 - \lambda_k} x_i \in A$ .

$\Rightarrow x \in A$ . Entonces hemos cogido  $x \in C_0(A)$  y llegado a  $x \in A \forall x \in C_0(A)$  si  $A$  convexo.  
 $\Rightarrow A = C_0(A)$ .

2)

demostración quitada ■

**Lema 1.2.2**

$x_1, \dots, x_k \in \mathbb{R}^n$  son afínmente dependientes si y sólo si existen  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$  no todos nulos tales que  $\sum_{i=1}^k \lambda_i = 0$  y  $\sum_{i=1}^k \lambda_i x_i = 0$ .

linealmente dependientes si quitamos  $\sum \lambda = 0$ .

**Demostración 1.4**

demostración quitada ■

**Teorema 1.2.1 — de Caratheodory.**

Sea  $S \subseteq \mathbb{R}^n$ , si  $x \in C_0(S)$  entonces puede ser expresado como combinación convexa estricta de a lo sumo  $n + 1$  puntos de  $S$  afínmente independientes.

**Demstración 1.5**

Sea  $x \in C_0(S)$ , de la forma  $x = \sum_{i=1}^k \lambda_i x_i$  para ciertos  $x_1, \dots, x_k \in A$ , con  $\lambda_1, \dots, \lambda_k \geq 0$ ,  $\sum_{i=1}^k \lambda_i = 1$  (una combinación convexa).

Elegimos esta combinación la más corta posible ( $k$  más pequeña posible), con  $\lambda_i > 0 \forall i$ .

Por reducción al absurdo, supongamos que  $x_1, \dots, x_k$  son afínmente dependientes. Por el Lema 1.2.2, existen  $\alpha_1, \dots, \alpha_k \in \mathbb{R}$  no todos nulos tales que  $\sum_{i=1}^k \alpha_i = 0$  y  $\sum_{i=1}^k \alpha_i x_i = 0$ .

Ordenando de los índices, podemos suponer que

$$\frac{\lambda_k}{\alpha_k} = \min \left\{ \frac{\lambda_i}{\alpha_i} \mid \alpha_i > 0 \right\}.$$

Si podemos escribir  $x$  como combinación lineal convexa de  $x_1, \dots, x_{k-1}$  tendremos una contradicción (se había supuesto la combinación original la más corta posible).

$$\begin{aligned} x &= \sum_{i=1}^k \lambda_i x_i = \sum_{i=1}^{k-1} \lambda_i x_i + \lambda_k x_k \\ &= \sum_{i=1}^{k-1} \lambda_i x_i - \sum_{i=1}^{k-1} \frac{\lambda_k}{\alpha_k} \alpha_i x_i + \sum_{i=1}^{k-1} \frac{\lambda_k}{\alpha_k} \alpha_i x_i + \lambda_k x_k \\ &= \sum_{i=1}^{k-1} \left( \lambda_i - \frac{\lambda_k}{\alpha_k} \alpha_i \right) x_i + \sum_{i=1}^{k-1} \frac{\lambda_k}{\alpha_k} \alpha_i x_i + \lambda_k x_k. \end{aligned}$$

Los dos últimos términos suman a cero, ya que

$$\sum_{i=1}^{k-1} \frac{\lambda_k}{\alpha_k} \alpha_i x_i + \lambda_k x_k = \frac{\lambda_k}{\alpha_k} \sum_{i=1}^{k-1} \alpha_i x_i + \frac{\lambda_k}{\alpha_k} \alpha_k x_k = \frac{\lambda_k}{\alpha_k} \sum_{i=1}^k \alpha_i x_i = \frac{\lambda_k}{\alpha_k} \cdot 0 = 0.$$

Por tanto, sólo nos queda ver que el primer término es una combinación lineal convexa. Para demostrar, los coeficientes tienen que sumar a 1:

$$\sum_{i=1}^{k-1} \left( \lambda_i - \frac{\lambda_k}{\alpha_k} \alpha_i \right) = \left( \sum_{i=1}^{k-1} \lambda_i + \lambda_k \right) - \left( \lambda_k + \sum_{i=1}^{k-1} \frac{\lambda_k}{\alpha_k} \alpha_i \right) = 1 - \frac{\lambda_k}{\alpha_k} \sum_{i=1}^k \alpha_i = 1 - \frac{\lambda_k}{\alpha_k} \cdot 0 = 1,$$

y hay que demostrar que cada coeficiente  $\lambda_i - \frac{\lambda_k}{\alpha_k} \alpha_i$  es no negativo. Usando el orden asumido de los índices  $\frac{\lambda_k}{\alpha_k} = \min \left\{ \frac{\lambda_i}{\alpha_i} \mid \alpha_i > 0 \right\}$  .:

- Si  $\alpha_i > 0$ , entonces  $\frac{\lambda_k}{\alpha_k} \leq \frac{\lambda_i}{\alpha_i}$  luego  $\lambda_i - \frac{\lambda_k}{\alpha_k} \alpha_i \geq 0$ .
- Si  $\alpha_i = 0$ , entonces  $\lambda_i - \frac{\lambda_k}{\alpha_k} \alpha_i = \lambda_i \geq 0$ .
- Si  $\alpha_i < 0$ , entonces  $\alpha_k \geq 0$  porque  $\sum \alpha = 0$  y también  $\lambda_k \geq 0 \Rightarrow \frac{\lambda_k}{\alpha_k} \geq 0 \geq \frac{\lambda_i}{\alpha_i}$  luego  $\frac{\lambda_k}{\alpha_k} \alpha_i \leq \lambda_i \Rightarrow \lambda_i - \frac{\lambda_k}{\alpha_k} \alpha_i \geq 0$ .

Así queda demostrado que un punto  $x \in C_0(S)$  puede ser expresado como una combinación convexa de  $k-1$  puntos de  $S$  independientes. Lo cuál es una contradicción de la suposición de ser la combinación más corta.

Si elegimos  $k \leq n+1$ , ya hemos terminado la demostración. Si elegimos  $k > n+1$ , solo hace falta repetir el mismo proceso en bucle hasta que  $k \leq n+1$ . ■

**■ Ejemplo 1.11**

Consideremos el conjunto  $S \subset \mathbb{R}^2$ , donde  $S = \{(0,0), (0,1), (1,0), (1,1)\}$ .

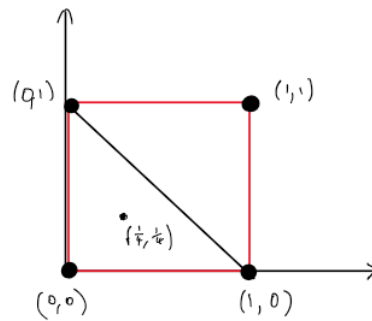


Figura 1.13: Teorema de Caratheodory

El teorema de Caratheodory nos dice que un punto, como por ejemplo  $(\frac{1}{4}, \frac{1}{4})$ , puede ser expresado por una combinación convexa de como mucho 3 puntos de  $S$ . Vemos que la combinación convexa  $\frac{1}{2} \langle 0, 0 \rangle + \frac{1}{4} \langle 0, 1 \rangle + \frac{1}{4} \langle 1, 0 \rangle = \langle \frac{1}{4}, \frac{1}{4} \rangle$ . ■

### ■ Ejemplo 1.12

Una aplicación del teorema de Caratheodory: Demostrar que la envolvente convexa de un conjunto compacto es compacta.

Sea  $X$  compacto. Tomamos una sucesión en  $C_0(X)$  y demostramos que tiene una sub-sucesión convergente cuyo punto de acumulación pertenece a  $C_0(X)$ .

Por Caratheodory, una sucesión en  $C_0(X)$  se puede expresar como  $\{\sum_{i=1}^{n+1} \alpha_i^k x_i^k\}$ , donde  $\forall i, k \alpha_i^k \geq 0, x_i^k \in X$  y  $\sum_{i=1}^{n+1} \alpha_i^k x_i^k = 1$ .

Entonces la sucesión

$$\{(\alpha_1^k, \dots, \alpha_{n+1}^k), (x_1^k, \dots, x_{n+1}^k)\}$$

está acotada, y tiene punto de acumulación

$$\{(\alpha_1, \dots, \alpha_{n+1}), (x_1, \dots, x_{n+1})\}$$

que satisface  $\sum_{i=1}^{n+1} \alpha_i = 1$  y  $\alpha_i \geq 0, x_i \in X \forall i$ .

Por definición el vector  $\sum_{i=1}^{n+1} \alpha_i x_i$  pertenece a  $C_0(X)$  y hemos demostrado que es un punto de acumulación de  $\{\sum_{i=1}^{n+1} \alpha_i^k x_i^k\}$ . Entonces para cualquier subsucesión su punto de acumulación está dentro de  $C_0(X)$ , por lo tanto es cerrado.

Por ser un conjunto acotado y cerrado, que  $C_0(X)$  es compacta.

(Un conjunto es cerrado si incluye todos sus puntos de acumulación.) ■

## 1.3 Separación

En esta sección, consideremos los resultados más importantes sobre la definición y existencia de hiperplanos que separan conjuntos convexos. Veremos más adelante que los teoremas definidos en esta sección nos permiten establecer la existencia de los multiplicadores en los problemas de programación matemática con restricciones.

### Definición 1.3.1 — Hiperplano.

Un **hiperplano** en  $\mathbb{R}^n$  es un conjunto de la forma

$$H = \{x : p^T x = \alpha\} \text{ con } p \in \mathbb{R}^n, p \neq 0 \text{ y } \alpha \in \mathbb{R}$$

O de forma equivalente, si  $x_0 \in H$   $H = \{x : p^T(x - x_0) = 0\}$ , donde  $x_0$  es un vector en el hiperplano.

A  $p$  se le llama **vector normal** al hiperplano.

#### ■ Ejemplo 1.13

Un hiperplano en  $\mathbb{R}^2$  es la línea  $(1, 1)^T(x, y) = 4$ , es decir  $x + y = 4$ ,

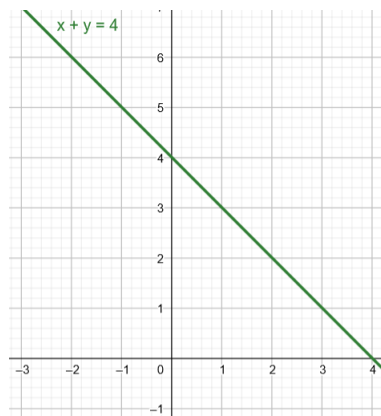


Figura 1.14: Hiperplano en 2D

#### ■ Ejemplo 1.14

Un hiperplano en  $\mathbb{R}^3$ ,  $(2, 2, -1)^T(x, y, z) = 0$  es el plano  $z = 2x + 2y$

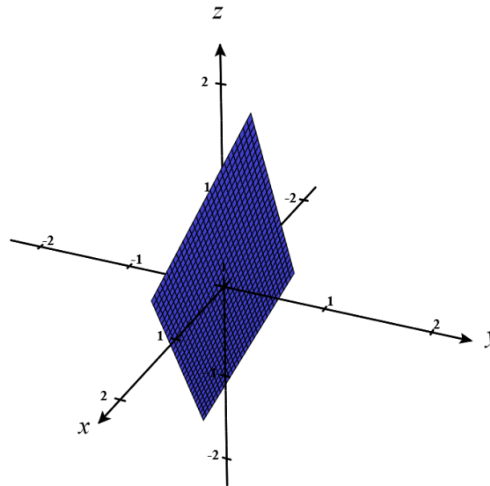


Figura 1.15: Hiperplano en 3D

#### Definición 1.3.2 — Semiespacio.

Un hiperplano define dos **semiespacios cerrados**,  $H^- = \{x : p^T x \leq \alpha\}$  y  $H^+ = \{x : p^T x \geq \alpha\}$  y dos **semiespacios abiertos**,  $\{x : p^T x < \alpha\}$  y  $\{x : p^T x > \alpha\}$ .

#### ■ Ejemplo 1.15

El hiperplano  $(1, 1)^T(x, y) = 4$ , es decir  $x + y = 4$ , define los semiespacios cerrados  $H^- = \{(1, 1)^T(x, y) \leq 4\}$

$4\}$ ,  $H^+ = \{(1, 1)^T(x, y) \leq 4\}$  y los semiespacios abiertos  $H^- = \{(1, 1)^T(x, y) < 4\}$ ,  $H^+ = \{(1, 1)^T(x, y) > 4\}$

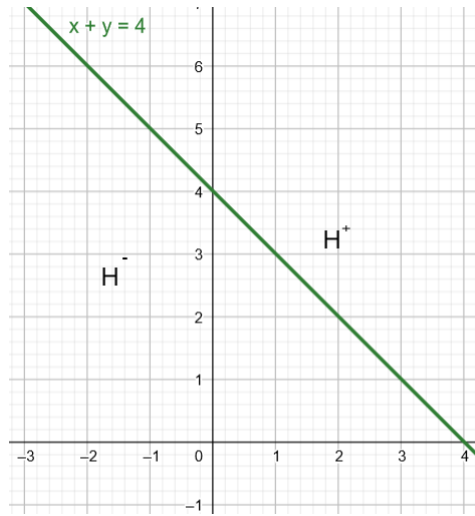


Figura 1.16: Semiespacios en 2D

### Proposición 1.3.1

Todo semiespacio es un conjunto convexo.

### Demostración 1.6

Consideremos  $a \in \mathbb{R}^n$ ,  $\alpha \in \mathbb{R}$ , que definen el semiespacio

$$S = \{x \in \mathbb{R}^n \mid a^T x \leq \alpha\}.$$

Sean  $x, y \in S$ ,  $t \in [0, 1]$ , entonces

$$\begin{aligned} a^T (tx + (1-t)y) &= ta^T x + (1-t)a^T y \\ &\leq t\alpha + (1-t)\alpha \\ &\leq \alpha. \end{aligned}$$

Luego  $tx + (1-t)y \in S$  y, por lo tanto,  $S$  es convexo.

### Definición 1.3.3

Sean  $S_1, S_2 \subseteq \mathbb{R}^n$  no vacíos. Diremos que  $H = \{x : p^T x = \alpha\}$

–**Separa**  $S_1$  y  $S_2$  si  $p^T x_1 \geq \alpha$  y  $p^T x_2 \leq \alpha$ ,  $\forall x_1 \in S_1$  y  $x_2 \in S_2$ ,

–**Separa estrictamente**  $S_1$  y  $S_2$  si  $p^T x_1 > \alpha$  y  $p^T x_2 < \alpha$ ,  $\forall x_1 \in S_1$  y  $x_2 \in S_2$ ,

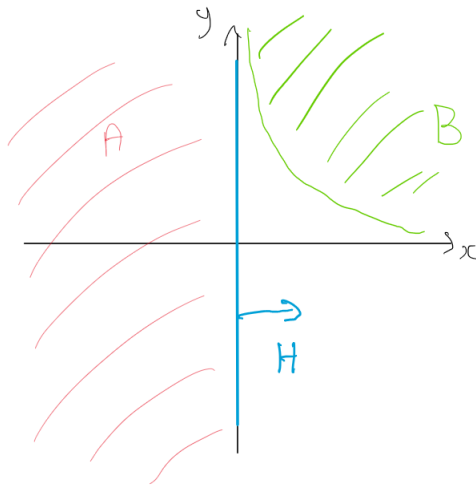
–**Separa fuertemente**  $S_1$  y  $S_2$  si  $p^T x_1 \geq \alpha + \varepsilon$  y  $p^T x_2 \leq \alpha$ ,  $\forall x_1 \in S_1$  y  $x_2 \in S_2$  con  $\varepsilon > 0$ .

### Ejemplo 1.16

- El hiperplano  $\{x \in \mathbb{R}^2 \mid x_2 - x_1 = 1\}$  separa fuertemente los conjuntos  $T = \{x \in \mathbb{R}^2 \mid 0 \leq x_1 \leq 1, 3 \leq x_2 \leq 5\}$  y  $S = \{x \in \mathbb{R}^2 \mid x_2 = 0, x_1 \geq -1\}$ ,
- El hiperplano  $\{x \in \mathbb{R} \mid x = 1\}$  separa los conjuntos  $T = \{x \in \mathbb{R} \mid x \leq 1\}$  y  $S = \{x \in \mathbb{R} \mid x \geq 1\}$ ,
- El hiperplano  $\{x \in \mathbb{R}^2 \mid x_1 = 0\}$  separa los conjuntos  $T = \{x \in \mathbb{R}^2 \mid x_1 < 0, x_2 \geq \frac{-1}{x_1}\}$  y  $S =$

- $\{x \in \mathbb{R}^2 | x_1 > 0, x_2 > \frac{1}{x_1}\}$ ,
- El hiperplano  $\{x \in \mathbb{R}^2 | x_2 - x_1 = 1\}$  no separa los conjuntos  $T = \{x \in \mathbb{R}^2 | x_2 \geq 1\}$  y  $S = \{x \in \mathbb{R}^2 | x_2 = 0\}$ ,
- El hiperplano  $\{x \in \mathbb{R}^2 | x_2 = 0\}$  no separa los conjuntos  $T = \{x \in \mathbb{R}^2 | x_2 = 0, x_1 \leq -1\}$  y  $S = \{x \in \mathbb{R}^2 | x_2 = 0, x_1 \geq 1\}$ .

■ **Ejemplo 1.17**



Sean los conjuntos  $A = \{(x, y) \in \mathbb{R}^2 | x \leq 0\}$  y  $B = \{(x, y) \in \mathbb{R}^2 | x > 0, y \geq \frac{1}{x}\}$ .

A y B son conjuntos no vacíos y  $A \cap B = \emptyset$ , pero no se pueden separar estrictamente. Solo podemos definir el hiperplano  $H = \{(x, y) \in \mathbb{R}^2 | x = 0\}$  que los separa.

Figura 1.17: Separación

**Definición 1.3.4**

Sea  $S \subseteq \mathbb{R}^n$  no vacío y  $\bar{x} \in \partial S$ . Diremos que  $H = \{x : p^T x = \alpha\}$  es un **hiperplano soporte** de S en  $\bar{x}$  si  $S \subseteq H^+$  o  $S \subseteq H^-$ .

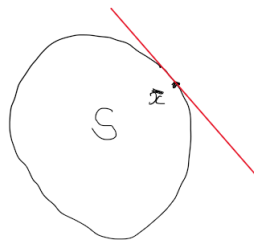


Figura 1.18: Hiperplano Soporte

■ **Ejemplo 1.18**

El hiperplano  $x_1 = 1$  en  $\mathbb{R}^2$  es un hiperplano soporte de la bola Euclídea  $\{x | |x| \leq 1\}$  en el punto  $x = (1, 0)$ .

**Teorema 1.3.1 — de la proyección.**

Sea  $S \subseteq \mathbb{R}^n$ , no vacío, cerrado y convexo e  $y \notin S$ . Entonces existe un único  $\bar{x} \in S$  que minimiza la distancia de y a S:  $\|y - \bar{x}\| \leq \|y - x\| \quad \forall x \in S$ .

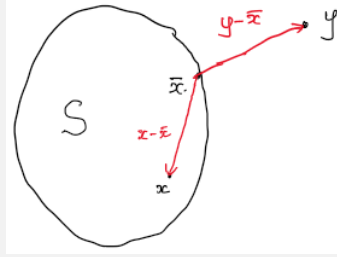


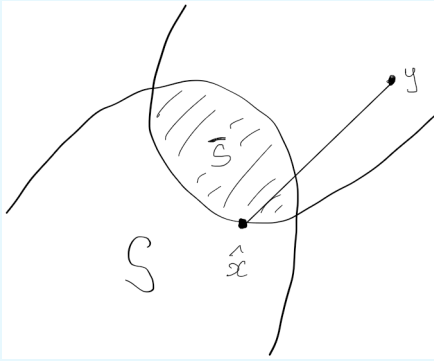
Figura 1.19: Proyección

Además,  $\bar{x}$  es el punto que minimiza la distancia sí y solo sí  $(y - \bar{x})^T(x - \bar{x}) \leq 0, \forall x \in S$ .

**Nota**  $\|x\| = \sqrt{\sum_i x_i^2}$  norma euclídea.

### Demostración 1.7

**Existencia:** Sea  $S$  no vacío,  $y \notin S$  y  $\hat{x} \in S$ . Definimos  $\bar{S} = S \cap \{x \in S \mid \|x - y\| \leq \|\hat{x} - y\|\}$ . Por tanto,  $\inf\{\|y - x\| \mid x \in S\} = \inf\|y - x\| \mid x \in \bar{S}$ .



$\bar{S}$  es convexo, cerrado y acotado (compacto) y, por el teorema de Weierstrass, el ínfimo se alcanza. Es decir,  $\exists \bar{x} \mid \|\bar{x} - y\| \leq \|x - y\| \forall x \in S$ .

Figura 1.20: Proyección

**Unicidad:** Supongamos  $\exists \bar{x}, \tilde{x}$  donde  $\bar{x} \neq \tilde{x} \mid \|y - \bar{x}\| = \|y - \tilde{x}\| = \gamma$ . Si consideramos el punto en el medio y su distancia de  $y$ , tenemos:

$$\|y - \frac{\bar{x} + \tilde{x}}{2}\| = \left\| \frac{y - \bar{x}}{2} + \frac{y - \tilde{x}}{2} \right\| \leq \frac{1}{2}\|y - \bar{x}\| + \frac{1}{2}\|y - \tilde{x}\| = \frac{1}{2}\gamma + \frac{1}{2}\gamma = \gamma.$$

donde  $\frac{\bar{x} + \tilde{x}}{2} \in S$ . Todas las desigualdades son igualdades porque si no, habríamos encontrado un valor más cercano a  $y$ . Si  $\frac{1}{2}\|y - \bar{x}\| + \frac{1}{2}\|y - \tilde{x}\| = \gamma$ , queremos decir que los vectores entre  $y$  y ambos puntos son diferentes por un escalar, es decir  $\exists \alpha : y - \bar{x} = \alpha(y - \tilde{x})$ .

Entonces su distancia es:

$$\gamma = \|y - \bar{x}\| = \|\alpha(y - \tilde{x})\| \Rightarrow |\alpha| = 1$$

Si  $\alpha = -1$ , tenemos  $y - \bar{x} = -(y - \tilde{x}) \mid y = \frac{\bar{x} + \tilde{x}}{2}$ . Esto es una contradicción, porque  $y \notin S$  y  $\frac{\bar{x} + \tilde{x}}{2} \in S$ . Entonces hay que tener  $\alpha = 1$ :

$$y - \bar{x} = y - \tilde{x} \Rightarrow \bar{x} = \tilde{x}.$$

llegando a la conclusión que el mínimo es único.

**Caracterización:** Suponemos que cumple  $(y - \bar{x})^T(x - \bar{x}) \leq 0 \forall x \in S$ . Entonces,

$$\|y - x\|^2 = \|y - \bar{x} + \bar{x} - x\|^2 = \|y - \bar{x}\|^2 + \|\bar{x} - x\|^2 + 2(y - \bar{x}, \bar{x} - x)$$

por la regla del coseno.

Tenemos  $2(y - \bar{x}, \bar{x} - x) = (y - \bar{x})(\bar{x} - x) \geq 0$  por lo que hemos supuesto. Porque  $\bar{x} - x = -(x - \bar{x})$ .

Entonces  $\|y - x\|^2 \geq \|y - \bar{x}\|^2 \forall x \in S \Rightarrow \bar{x}$  minimiza la distancia a  $y$ .

Ahora suponemos que  $\bar{x}$  es el mínimo. Consideremos  $x \in S$  arbitrario, y escribimos la combinación convexa como  $\bar{x} + \lambda(x - \bar{x}) \in S$ ,  $\lambda \in [0, 1]$ . La distancia de  $y$  a este punto es

$$\|y - \bar{x} - \lambda(x - \bar{x})\| \geq \|y - \bar{x}\|$$

porque  $\bar{x}$  es mínimo.

Vemos que

$$\begin{aligned} \|y - \bar{x}\|^2 &\leq \|y - \bar{x} - \lambda(x - \bar{x})\|^2 \\ \|y - \bar{x}\|^2 &\leq \|y - \bar{x}\|^2 + \|\lambda(x - \bar{x})\|^2 - 2\lambda(y - \bar{x})^T(x - \bar{x}) \\ 0 &\leq \lambda^2\|(x - \bar{x})\|^2 - 2\lambda(y - \bar{x})^T(x - \bar{x}) \\ 2(y - \bar{x})^T(x - \bar{x}) &\leq \lambda\|(x - \bar{x})\|^2 \end{aligned}$$

$\lambda$  es la medida de la distancia entre  $x$  y  $y - \bar{x}$ , entonces cuando tomamos  $\lambda \rightarrow 0$ , tenemos  $\bar{x}$  que es el punto de interés:

$$(y - \bar{x})^T(x - \bar{x}) \leq 0.$$

### Teorema 1.3.2

Sea  $S \subseteq \mathbb{R}^n$ , no vacío, cerrado y convexo e  $y \notin S$ . Entonces existe  $p \in \mathbb{R}^n, p \neq 0$  y existe  $\alpha \in \mathbb{R}$  tal que

$$p^T y > \alpha \quad \text{y} \quad p^T x \leq \alpha \quad \forall x \in S.$$

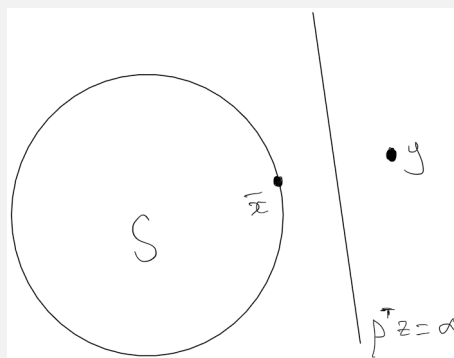


Figura 1.21: Separación

### Demostración 1.8

Por la teorema de proyección, sabemos que existe un único mínimo  $\bar{x} \in S$  caracterizado por

$$(y - \bar{x})^T(x - \bar{x}) \leq 0 \quad \forall x \in S.$$

Sea  $p = y - \bar{x} \neq 0$ . Entonces tenemos,

$$p^T(x - \bar{x}) \leq 0 \Rightarrow p^T x \leq p^T \bar{x} = \alpha,$$

donde podemos suponer que  $\alpha = p^T \bar{x} = \max\{p^T x | x \in S\}$ . Tenemos

$$\begin{aligned} (y - \bar{x})^T(x - \bar{x}) &\leq 0 \quad \forall x \in S \\ (y - \bar{x})^T x - (y - \bar{x})^T \bar{x} &\leq 0 \quad \forall x \in S \\ p^T x - \alpha &\leq 0 \quad \forall x \in S \end{aligned}$$

Si  $y \notin S, y \neq \bar{x}$ , sabemos que  $0 < \|y - \bar{x}\|^2$ . Entonces,

$$0 < \|y - \bar{x}\|^2 = (y - \bar{x})^T(y - \bar{x}) = p^T y - p^T \bar{x} \Rightarrow \alpha < p^T y.$$

### Corolario 1.3.1

Sea  $S \subseteq \mathbb{R}^n$ , no vacío, cerrado y convexo, entonces  $S$  es igual a la intersección de todos los semiespacios cerrados que lo contienen.

### Corolario 1.3.2

Sea  $S \subseteq \mathbb{R}^n$ , no vacío, convexo e  $y \notin cl(S)$ . Entonces existe un hiperplano que separa fuertemente  $S$  e  $y$ .

### Definición 1.3.5 — Cono.

Un conjunto  $K \subseteq \mathbb{R}^n$  se le denomina un **cono** si  $x \in K \Rightarrow \alpha x \in K$  por cualquier escalar  $\alpha \geq 0$ .

### Definición 1.3.6 — Envoltente Cónica.

Sea  $S$  un conjunto, la **envoltente cónica**,  $cono(S)$ , es el conjunto de todas las combinaciones cónicas de los puntos en  $S$ .

$$cono(S) = \left\{ \sum_{i=1}^n \alpha_i x_i \mid \alpha_i \geq 0, x_i \in S \right\}.$$



Figura 1.22: Envoltente Cónica

El conjunto factible de un problema de optimización suele escribirse como  $\{x | x \geq 0, Ax = b\}$ . Queremos saber si ese conjunto es vacío o no.

### Teorema 1.3.3 — de Farkas.

Sea  $A \in \mathbb{R}^{m \times n}$  y  $b \in \mathbb{R}^m$ . Entonces, exactamente una de las siguientes afirmaciones es verdadera:

1.  $Ax = b, x \geq 0, x \in \mathbb{R}^n$
  2.  $A^T y \geq 0, b^T y < 0, y \in \mathbb{R}^m$
- donde  $x \geq 0$  significa que todos los elementos de  $x$  son no negativos.

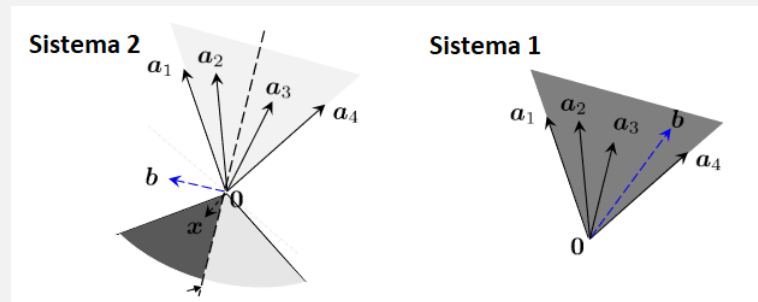


Figura 1.23: Teorema de Farkas

El teorema (o lema) de Farkas nos dice si un sistema de desigualdades lineales tiene solución, o no. Nos dice que existe una solución (conjunto no vacío), o existe  $y$  tal que un hiperplano que le separe del conjunto.

#### Interpretación geométrica

Sea  $C(A)$  la envolvente cónica de las columnas  $\tilde{a}_i$  de  $A \in \mathbb{R}^{m \times n}$ ,  $C(A) = \{Ax | x \geq 0\}$ . Se ve que  $C(A)$  es el conjunto de vectores  $b$  para la primera condición del teorema, y que el vector  $y$  es ortogonal al hiperplano que separa  $b$  y  $C(A)$ . Entonces,  $b$  solo puede pertenecer a  $C(A)$  si no existe un hiperplano que le separe de  $C(A)$ .

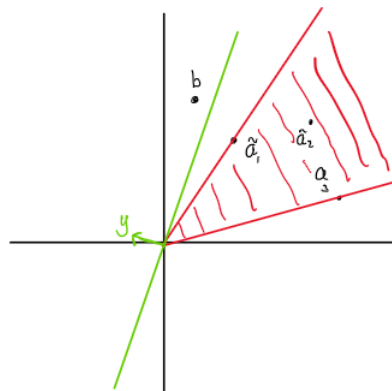


Figura 1.24: Teorema de Farkas

#### Demostración 1.9

Suponemos que la primera condición tiene solución. Entonces  $Ax = b, x \geq 0$ . Tomemos  $y \in \mathbb{R}^m$ , y multiplicamos:  $y^T Ax = y^T b$ . Si  $y^T A \geq 0 \Rightarrow y^T Ax \geq 0 \Rightarrow y^T b \geq 0$ .

Por tanto, la segunda condición no tiene solución.

Suponemos que la primera condición no tiene solución, y consideremos el conjunto convexo y cerrado  $S = \{v \in \mathbb{R}^m | v = Ax, x \geq 0\}$  (cono).  $b \notin S$ , porque la primera condición no tiene solución. Entonces  $\exists p \in \mathbb{R}^m, \alpha \in \mathbb{R}$  tales que  $p^T b > \alpha, p^T v \leq \alpha \forall v \in S$ .

Observamos que  $0 \in S$ , pues  $0 = A \cdot 0$ , luego  $p^T 0 \leq \alpha \Rightarrow 0 \leq \alpha$ . Por tanto,  $p^T b > \alpha \geq 0 \Rightarrow p^T b > 0$ .

Si probamos que  $p^T A \leq 0$ ,  $p$  sería solución de la segunda condición.

Tomamos cualquier  $v \in S$ , entonces  $v = Ax, x \geq 0$ . Tenemos  $p^T v \leq \alpha \Rightarrow p^T Ax \leq \alpha, x \geq 0$ . Si el vector  $p^T A$  tuviese una coordenada positiva, por ejemplo  $(p^T A)_i > 0$ , tomamos  $(0, \dots, 0, \lambda, 0, \dots, 0) = x(\lambda)$ . Evaluamos  $p^T Ax$  en ese vector  $p^T Ax(\lambda) = (p^T A)_i \lambda$ . Si dejamos  $\lambda \rightarrow \infty$ , vemos que  $p^T Ax \rightarrow \infty$ , lo que contradice que  $p^T Ax \leq \alpha \forall x \geq 0$ . Entonces  $p^T A \leq 0$  y  $p$  es solución de la segunda condición. ■

### ■ Ejemplo 1.19

Sean

$$A = \begin{pmatrix} 6 & 4 \\ 3 & 0 \end{pmatrix} \quad b_1 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

Tenemos:

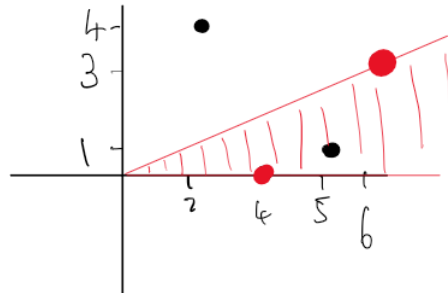


Figura 1.25: Ejemplo teorema de Farkas

Para  $b_1$  el teorema dice que:

- 1 Existen  $x \geq 0$  tales que  $6x_1 + 4x_2 = 5$  y  $3x_1 = 1$ , o,
- 2 Existen  $y$  tales que  $6y_1 + 3y_2 \geq 0, 4y_1 \geq 0$ , y  $5y_1 + y_2 < 0$ .

Desde 1) vemos que  $x_1 = \frac{1}{3}$  y  $x_2 = \frac{3}{4}$ . Ambos son positivos, entonces por el teorema de Farkas, existe una solución al sistema  $Ax = b$ . Podemos comprobar que no existe un  $y \in \mathbb{R}^m$  para 2), donde  $6y_1 + 3y_2 \geq 0 \Rightarrow -2y_1 \leq y_2$ , en  $5y_1 + y_2 < 0 \Rightarrow y_1 < 0$ , lo cual es una contradicción.

Para  $b_2$  el teorema dice que:

- 1 Existen  $x \geq 0$  tales que  $6x_1 + 4x_2 = 2$  y  $3x_1 = 4$ , o,
- 2 Existen  $y$  tales que  $6y_1 + 3y_2 \geq 0, 4y_1 \geq 0$ , y  $2y_1 + 4y_2 < 0$ .

Desde 1) vemos que  $x_1 = \frac{4}{3}$  y  $x_2 = -3$ . Por lo tanto, este punto no cumple la primera condición del teorema de Farkas. Comprobamos este resultado en la segunda condición:  $6x_1 + 4x_2 = 2 \Rightarrow -2y_1 \leq y_2$ , en  $2y_1 + 4y_2 < 0 \Rightarrow -6y_1 < 0, y_1 \geq 0$ , entonces existe un hiperplano que separa el punto  $b_2$  y la envolvente convexa de  $A$ .

### 1.3.1 Variantes de Farkas

Hay variantes del lemma de Farkas, que se puede aplicar a diferentes situaciones. Cada variante se desarrolla desde otro variante. Vemos el desarrollo de un variante:

Original

1.  $Ax = b, x \geq 0, x \in \mathbb{R}^n$
2.  $A^T y \geq 0, b^T y < 0, y \in \mathbb{R}^m$

Variante

1.  $Ax \leq b, x \geq 0, x \in \mathbb{R}^n$
2.  $A^T y \geq 0, b^T y < 0, y \geq 0 \in \mathbb{R}^m$

Para demostrar que el variante es cierto (e igual al original) intentaremos transformarlo al original,

y demostrar que cumple las condiciones del original.

Si tenemos  $A\vec{x} \leq \vec{b}$  con  $\vec{x} \geq 0$ , podemos sumar una variable de holgura positiva para tener igualdad.  $A\vec{x} + I\vec{s} = \vec{b}$ , donde  $I$  es la matriz de identidad.

Concatenamos horizontalmente las variable en matrices  $(A|I)\begin{pmatrix} x \\ s \end{pmatrix} = \vec{b}$  y si  $x \geq 0$  y  $s \geq 0$ , podemos decir que  $\begin{pmatrix} x \\ s \end{pmatrix} \geq 0$  concatenando verticalmente. Dejo que  $A' = (A|I)$  y  $x' = \begin{pmatrix} x \\ s \end{pmatrix}$ , entonces tengo  $A'x' = \vec{b}$ ,  $x' \geq 0$ . Lo cual es sistema 1 del original. Concluimos, que si 1 del variante es cierto, también lo es 1 del original.

Si es cierto, entonces desde el original, 2 es falso  $\nexists y : A^T y \geq 0, b^T y < 0$ . Cambiamos esto a la forma del variante 2. Tenemos  $b^T y < 0$ , cuando  $A^T y \geq 0$  y  $y \geq 0$ . Podemos decir por ser ambos positivos  $\frac{A^T y}{y} \geq 0$ . Lo cual es igual a  $\left(\frac{A^T}{I}\right)\vec{y} \geq 0$ .

Si  $A' = (A|I)$ , entonces  $A'^T = \begin{pmatrix} A^T \\ I \end{pmatrix}$ . Substituyendo arriba, tenemos  $A'^T \vec{y} \geq 0$  y  $b^T y < 0$ . Entonces  $A'^T \vec{y} \geq 0$  implica  $A^T \vec{y} \geq 0$ , y tenemos  $b^T y < 0$ . Pero ya hemos demostrado que  $\nexists y : A^T y \geq 0, b^T y < 0$ , entonces 2 del original también es falso si 1 del variante es cierto.

Pasos del razonamiento:

- 1 del variante cierto implica 1 del original cierto
- 1 del original cierto implica 2 del original falso
- 2 del original falso implica 2 del variante falso.

#### Teorema 1.3.4 — de Gordan.

Sea  $A \in \mathcal{M}^{m \times n}$ . Uno y solo uno de los siguientes sistemas tiene solución:

1.  $Ax < 0$  para algún  $x \in \mathbb{R}^n$ , o
2.  $A^T y = 0, y \geq 0$  para algún  $y \in \mathbb{R}^m$  ( $y \neq 0$ ).

#### Teorema 1.3.5 — de Tucker.

Si  $A \in \mathcal{M}^{x \times x}$  es anti-simétrica entonces existe  $w \in \mathbb{R}^n, w \geq 0$  tal que  $Aw \geq 0$  y además  $Aw + w > 0$ .

## 1.4 Problemas

- Una compañía petrolífera debe determinar cuántos barriles de petróleo hay que extraer en los próximos dos años. Si la compañía extrae  $x_1$  millones de barriles durante un año, se podrá vender cada barril a  $30 - x_1$  euros. Si extrae  $x_2$  millones de barriles durante el segundo año, se podrá vender cada barril a  $35 - x_2$  euros. El coste para extraer  $x_1$  millones de barriles en el primer año es  $x_1^2$  millones de euros y el coste para extraer  $x_2$  millones de barriles durante el segundo año es  $2x_2^2$  millones de euros. Se puede obtener como máximo un total de 20 millones de barriles, y se puede gastar como máximo 250 millones de euros en la extracción. Formule un problema de optimización para maximizar los ingresos de la compañía.
- Considere (dibuja) la región factible dada por las restricciones

$$1 - x_1^2 - x_2^2 \geq 0, \quad \sqrt{2} - x_1 - x_2 \geq 0, \quad x_2 \geq 0.$$

Para los siguientes puntos, determine si son factible o no, y si son factibles, si son puntos interiores o en la frontera de la región factible:  $x_a = (\frac{1}{2}, \frac{1}{2})^T$ ,  $x_b = (1, 0)^T$ ,  $x_c = (-1, 0)^T$ ,  $x_d = (-\frac{1}{2}, 0)^T$  y  $x_e = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ .

- Dibuje la región factible del problema de optimización

$$\begin{array}{ll} \text{maximizar} & f(x) = x_1 \\ \text{sujeto a} & (x_1 - 1)^2 + x_2^2 = 1 \\ & (x_1 + 1)^2 + x_2^2 = 1 \\ & x \geq 0. \end{array}$$

¿Existen óptimos locales o globales?

- Considerando el teorema de Weierstrass (Teorema 1.1.1), ¿qué podemos afirmar de las siguientes funciones definidas en  $\mathbb{R}^2$ ?
  - $f(x_1, x_2) = x_1 e^{x_1 + x_2}$ ,
  - $g(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2} & (x_1, x_2) \neq (0, 0) \\ 0 & (x_1, x_2) = (0, 0). \end{cases}$
- Obténganse geoméricamente los óptimos globales pedidos de los siguientes problemas:

$$\begin{array}{ll} \text{Min} & 3x_1 + 2x_2 \\ \text{Sujeto a} & x_1 + x_2 \leq 2 \\ & 3x_1 + x_2 \leq 3 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

b.

$$\begin{array}{ll} \text{Min} & 2x_1 + x_2 \\ \text{Sujeto a} & x_2 - e^{x_1} \geq 0 \end{array}$$

6. Dar un ejemplo de una función que no tiene un mínimo o máximo global.
7. Si  $A$  es un subconjunto convexo no vacío de  $\mathbb{R}^n$  y  $c \in \mathbb{R}$ , se define el producto de  $A$  por el escalar  $c$  como el conjunto

$$cA = \{z \in \mathbb{R}^n \mid z = cx, x \in A\}.$$

Pruébese que  $cA$  es un conjunto convexo.

8. Demuestra que los siguientes conjuntos son convexos
- $S = \{x \in \mathbb{R}^3 \mid x_1 + 2x_2 - x_3 = 2\}$
  - $S = \{(x, y) \in \mathbb{R}^2 \mid x + 2y \leq 6, 2x + y \leq 6\}$
  - $S = \{(x, y) \in \mathbb{R}^2 \mid |y| < 3, |x - 1| \leq 2\}$
9. Pruébese que la función  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) = a^T x + b$ ,  $a, x \in \mathbb{R}^n$ ,  $b \in \mathbb{R}$ , es cóncava y convexa a la vez.
10. Determine si las siguientes funciones son convexas, cóncavas, o ambas en  $\mathbb{R}$ . Si la función es cóncava o convexa, indique si es estrictamente o no.
- $f(x) = 3x^2 + 4x - 5$
  - $f(x) = \exp(x^2)$
  - $f(x) = 7x - 15$
  - $f(x) = \sqrt{1 + x^2}$
  - $f(x) = 4 - 5x + 3x^2$
  - $f(x) = 2x^4 + 3x^3 + 4x^2$
  - $f(x) = \frac{x}{1 + x^4}$

11. Determine si la función

$$f(x_1, x_2) = 2x_1^2 - 3x_1x_2 + 5x_2^2 - 2x_1 + 6x_2$$

es convexa, cóncava, ambas o ninguna con  $x \in \mathbb{R}^n$ .

12. ¿Se puede separar los siguientes conjuntos?  $S = \{x \in \mathbb{R}^2 \mid x_1 > 0, x_2 \geq \frac{1}{x_1}\}$ ,  $T = \{x \in \mathbb{R}^2 \mid x_1 < 0, x_2 \geq \frac{-1}{x_1}\}$ .





# Tema 2 - Optimización sin restricciones

<b>2</b>	<b>Optimización sin restricciones</b> . . . . .	<b>33</b>
2.1	Introducción	
2.2	Condiciones de Optimalidad	
2.3	Métodos Iterados	
2.4	Problemas	

part.1chapter.1section.1.1section.1.2section.1.3subsection.1.3.1

chapter.3section.3.1section.3.2subsection.3.2.1subsection.3.2.2s





## 2. Optimización sin restricciones

### 2.1 Introducción

En este tema, pasamos a considerar problemas de optimización **no lineales**, donde no incluimos *restricciones*, con la forma:

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeto a: } x \in S \end{aligned}$$

donde  $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ .

### 2.2 Condiciones de Optimalidad

#### Definición 2.2.1

Un punto  $x^* \in S$  es un **mínimo local o relativo** si existe  $\varepsilon > 0$  tal que  $f(x) \geq f(x^*)$  para todo  $x \in S$  tal que  $\|x - x^*\| < \varepsilon$ .

Diremos que  $x^*$  es un **mínimo global** si  $f(x) \geq f(x^*)$  para todo  $x \in S$ .

**Nota** Si las desigualdades son estrictas se dicen mínimos (locales o globales) estrictos.

El teorema de Weierstrass establece que si  $f$  es continua y  $S$  es compacto, existe una solución del problema de optimización. El objetivo de este tema es considerar métodos para hallar esta solución, y caracterizarla, para problema de optimización *sin restricciones*.

En problemas de optimización reales, e incluso problemas teóricos difíciles, no siempre es posible hallar una solución global. Como veremos más adelante con métodos iterativos, será común acabar con una solución local (y aproximada, en muchos casos). Para asegurar la optimalidad global será preciso que el problema verifique ciertas propiedades (convexidad).

#### Definición 2.2.2

Diremos que la función  $f$  es convexa en  $S$  convexo si para todo  $x_1, x_2 \in S$ ,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad \forall \lambda \in [0, 1]$$

Si la desigualdad es estricta para  $x_1 \neq x_2, \lambda \in (0, 1)$  diremos que  $f$  es estrictamente convexa. Diremos que  $f$  es (estrictamente) cóncava si  $-f$  es (estrictamente) convexa.

**Teorema 2.2.1 — Condición necesaria de optimalidad (de primer orden).**

Sea  $f$  de clase  $C^1$  en  $S$  abierto no vacío y  $x^*$  un mínimo local en  $S$ , entonces  $\nabla f(x^*) = 0$ .

**Nota** Esta condición es **necesaria** y no **suficiente**, porque un punto  $x^*$  con  $\nabla f(x^*) = 0$  podría ser un punto de silla, ni un máximo local ni un mínimo local.

**Demostración 2.1**

Supongamos por contradicción que  $\nabla f(x^*) \neq 0$ . Definimos el vector  $p = -\nabla f(x^*)$  y notamos que:

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$$

Dado que  $\nabla f$  es continuo cerca de  $x^*$ , existe un escalar  $T > 0$  tal que:

$$p^T \nabla f(x^* + tp) < 0 \quad \forall t \in [0, T]$$

Para algún  $\hat{t} \in [0, T]$ , tenemos (desde la serie de Taylor):

$$f(x^* + \hat{t}p) = f(x^*) + \hat{t}p^T \nabla f(x^*)$$

Por lo tanto  $f(x^* + \hat{t}p) < f(x^*)$  para todo  $\hat{t} \in (0, T]$ . Hemos encontrado una dirección,  $p$ , que se aleja de  $x^*$  y  $f$  decrece. Por lo cual,  $x^*$  no es un mínimo local.  $\Rightarrow$  contradicción. ■

■ **Ejemplo 2.1**

Consideramos las funciones  $f_1(x_1, x_2) = 2 - x_1^2 - x_2^2$ ,  $f_2(x_1, x_2) = x_1^2 + x_2^2$  y  $f_3(x_1, x_2) = x_1^2 - x_2^2$ . Por la condición necesaria de optimalidad, vemos que por el gradiente de cada función, el único punto crítico es  $(0, 0)$ .

Este punto es máximo de  $f_1$ , mínimo de  $f_2$  y punto de silla de  $f_3$ , como se puede observar en la representación gráfica y como más adelante comprobaremos analíticamente.

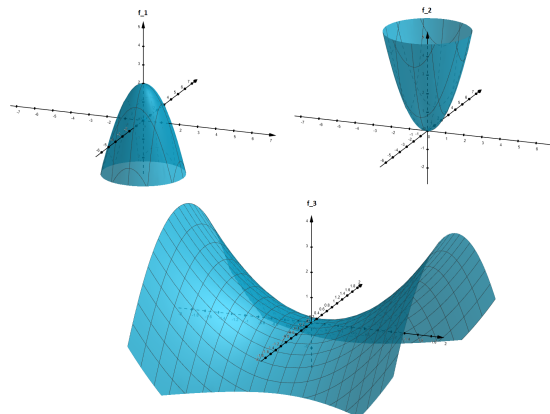


Figura 2.1: Representación de  $f_1, f_2, f_3$ .

Podemos distinguir mínimos (o máximos) locales usando las segundas derivadas de la función objetivo. ■

**Teorema 2.2.2 — Condición necesaria de optimalidad (de segundo orden).**

Sea  $f$  de clase  $C^2$  en  $S$  abierto no vacío y  $x^*$  un mínimo local en  $S$ , entonces  $\nabla^2 f(x^*)$  es semidefinida positiva.

**Nota** Si  $x^*$  es un máximo local, entonces  $\nabla^2 f(x^*)$  es semidefinida negativa.

**Demostración 2.2**

Sea  $x^*$  un mínimo local,  $d \in \mathbb{R}^n$  y  $\alpha \in \mathbb{R}$ . Consideremos el desarrollo de Taylor de segundo orden de  $f(x^* + \alpha d)$ ,

$$f(x^* + \alpha d) = f(x^*) + \alpha \nabla f(x^*)^T d + \frac{\alpha^2}{2} d^T \nabla^2 f(x^*) d + o(\alpha^2)$$

Recordando que  $\lim_{t \rightarrow 0} \frac{o(t)}{t} = 0$ ,  $\nabla f(x^*) = 0$  y que por ser  $x^*$  un mínimo  $f(x^* + \alpha d) - f(x^*) \geq 0$  se tiene que,

$$0 \leq d^T \nabla^2 f(x^*) d + \frac{o(\alpha^2)}{\alpha^2}$$

Tomando  $\alpha \rightarrow 0$ , se obtiene que  $\nabla^2 f(x^*)$  es semidefinida positiva. ■

■ **Ejemplo 2.2**

Ejemplo con un punto crítico y matriz Hessiana semidefinida positiva. Además,  $f(x, y) \geq 0 \forall x, y$  entonces es mínimo global.  $f(x, y) = (x - 1)^4 + (y - 1)^2$  ■

■ **Ejemplo 2.3**

Siguiendo con el ejemplo anterior, usamos la condición necesaria de optimalidad de segundo orden para identificar que tipo de puntos críticos encontramos para las funciones  $f_1(x_1, x_2) = 2 - x_1^2 - x_2^2$ ,  $f_2(x_1, x_2) = x_1^2 + x_2^2$  y  $f_3(x_1, x_2) = x_1^2 - x_2^2$ .

Calculando las hessianas:

$$Hf_1 = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix} \quad Hf_2 = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix} \quad Hf_3 = \begin{pmatrix} 2 & 2 \\ 0 & -2 \end{pmatrix}$$

Los valores propios:

$$\begin{aligned} \det(Hf_1 - \lambda I) &= (-2 - \lambda)^2 &\Rightarrow \lambda &= -2, \\ \det(Hf_2 - \lambda I) &= (2 - \lambda)^2 &\Rightarrow \lambda &= 2, \\ \det(Hf_3 - \lambda I) &= (-2 - \lambda)(2 - \lambda) &\Rightarrow \lambda &= -2 \text{ o } \lambda = 2. \end{aligned}$$

Vemos que la hessiana de  $f_1$  tiene autovalores negativos, siendo definida negativa y clasificando el punto crítico  $(0, 0)$  encontrado en el ejercicio anterior como un máximo. De la misma forma, vemos que  $(0, 0)$  para  $f_2$  es un mínimo y que para  $f_3$  es un punto de silla por tener autovalores de signos distintos (hessiana indefinida). ■

**Lema 2.2.1**

Sea  $f$  de clase  $C^1$  en  $S$  abierto, convexo y no vacío. Entonces  $f$  es convexa si y solo si

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) \quad \forall x_1, x_2 \in S$$

### Demostración 2.3

Sean  $x_1, x_2 \in S$  con  $x_1 < x_2$ , y  $c \in (x_1, x_2)$ . Aplicando el teorema del Valor Medio, tenemos

$$f(x_2) = f(x_1) + \nabla f(c)(x_2 - x_1)$$

Si  $f$  es convexa, entonces sabemos que  $\nabla f$  es creciente, y entonces  $\nabla f(c) \geq \nabla f(x_1)$ . Por lo tanto:

$$f(x_2) = f(x_1) + \nabla f(c)(x_2 - x_1) \geq f(x_1) + \nabla f(x_1)(x_2 - x_1)$$

### Teorema 2.2.3 — Condición suficiente de optimalidad.

Si  $f$  es convexa de clase  $C^1$  en  $S$  abierto, convexo, no vacío,  $x^* \in S$  tal que  $\nabla f(x^*) = 0$  entonces  $x^*$  es un mínimo global de  $f$  en  $S$ .

### Demostración 2.4

Por ser  $f$  convexa y de clase  $C^1$  en  $S$ , la definición de convexidad nos da  $f(x^* + h) - f(x^*) \geq \nabla f(x^*)h \quad \forall h \in \mathbb{R}^n$  tal que  $x^* + h \in S$ . Pero por ser  $x^*$  un punto crítico de  $f$ , tenemos  $\nabla f(x^*) = 0$ , dando  $f(x^* + h) \geq f(x^*) \quad \forall h \in \mathbb{R}^n$  tal que  $x^* + h \in S$ . Por lo tanto,  $x^*$  es un mínimo global. ■

### ■ Ejemplo 2.4

Consideramos la función convexa  $f(x_1, x_2) = (x_1 - 3)^2 + (x_2 + 1)^2$ . Evaluando  $\nabla f = 0$ , encontramos su único punto estacionario  $(3, -1)$ . Entonces, por la condición suficiente de optimalidad, sabemos que este punto es un mínimo global de la función. ■

### Lema 2.2.2

Sea  $f$  de clase  $C^2$  en  $S$  convexo, abierto, no vacío, entonces  $\nabla^2 f(x)$  es semidefinida positiva para todo  $x \in S$  si y solo si  $f$  es convexa en  $S$ .

### Teorema 2.2.4 — Condición suficiente de optimalidad.

Sea  $f$  de clase  $C^2$  y  $\nabla^2 f(x)$  semidefinida positiva en  $S$  abierto, convexo, no vacío,  $x^* \in S$  tal que  $\nabla f(x^*) = 0$ , entonces  $x^*$  es un mínimo global de  $f$  en  $S$ .

### Demostración 2.5

Sea  $x^*$  el mínimo de  $f$ . La serie de Taylor con termino de error sobre un punto  $x^* + p$  es:

$$f(x^* + p) = f(x^*) + \nabla f(x^*)^T p + \frac{1}{2} p^T \nabla^2 f(\xi) p.$$

Tenemos que  $\nabla f(x^*) = 0$ , entonces  $f(x^* + p) = f(x^*) + \frac{1}{2} p^T \nabla^2 f(\xi) p$ . Si  $\nabla^2 f(x)$  es continua y semidefinida positiva, también lo es  $\nabla^2 f(\xi)$ , para  $\|\xi - x^*\|$  suficientemente pequeño. Entonces tenemos  $f(x^* + p) > f(x^*) + \frac{1}{2} p^T \nabla^2 f(\xi) p$  por cualquier  $p$ , y  $f(x^*)$  es el mínimo global. ■

En casos donde la función de interés no sea convexa (o al menos no se puede demostrar), podemos

usar las mismas condiciones suficientes de optimalidad para hablar de óptimos locales.

**Teorema 2.2.5 — Condición suficiente de optimalidad (local).**

Sea  $f$  de clase  $C^2$  en  $S$  abierto no vacío. Supongamos que  $x^* \in S$  tal que  $\nabla f(x^*) = 0$  y  $\nabla^2 f(x^*)$  es definida positiva. Entonces  $x^*$  es un mínimo local estricto.

■ **Ejemplo 2.5**

Consideramos la función

$$f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2 - x_2 + 9$$

La condición necesaria de optimalidad nos da

$$\nabla f(x) = \begin{pmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{pmatrix} = 0$$

Vemos desde la segunda ecuación que  $x_2 = 1 - 2x_1$ , y sustituyendo en la primera nos da  $(x_1 - 1)(x_1 - 2) = 0$ . Por lo tanto, tenemos dos puntos críticos:

$$x_a = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad x_b = \begin{pmatrix} 2 \\ -3 \end{pmatrix}.$$

La Hessiana de la función es  $\nabla^2 f(x) = \begin{pmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{pmatrix}$ . Resolviendo la Hessiana en los dos puntos críticos:

$$\nabla^2 f(x_a) = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} \quad \nabla^2 f(x_b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}. \quad (2.1)$$

La hessiana en  $x_b$  es definida positiva, entonces  $x_b$  es un mínimo local de  $f(x)$ . La hessiana en  $x_a$  es indefinida, entonces  $x_a$  no es un mínimo o máximo local. Podemos ver que la función  $f(x)$  no tiene máximo o mínimo global, por no estar acotada cuando  $x_1 \rightarrow \pm\infty$ . ■

## 2.3 Métodos Iterados

En problemas reales, muchas veces, es imposible encontrar una solución analítica al problema. Existen métodos numéricos para tratar de aproximar las soluciones en estos casos.

*ejemplo para motivar?*

### 2.3.1 Método de Newton

Si  $f$  es dos veces diferenciable, se puede aproximar  $f$  localmente por una función cuadrática:

$$f(x) \approx q(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

Usando la condición necesaria de optimalidad  $\nabla f(x) = 0$ , tenemos  $0 = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k)$  y el método de Newton consiste en la iteración:

$$x_{k+1} = x_k - \frac{\nabla f(x_k)^T}{\nabla^2 f(x_k)}$$

donde  $-\frac{\nabla f(x_k)}{\nabla^2 f(x_k)}$  se llama la *dirección de Newton*.

Por la condición necesaria de segundo orden de mínimo, si  $x^*$  es un mínimo de  $f$  entonces  $\nabla^2 f(x^*)$  es semidefinida positiva. Si suponemos que  $f$  es de clase  $C^2$ , podemos suponer que  $\nabla^2 f(x_k)$  será definida positiva para  $x_k$  suficientemente cerca de  $x^*$  y por tanto  $x_{k+1}$  estará bien definido.

### ■ Ejemplo 2.6

Consideramos el problema de optimización

$$\begin{aligned} &\text{Minimizar } f(x) = 5x^5 + 2x^3 - 4x^2 - 3x + 2 \\ &\text{sujeto a } x \in \mathbb{R}^n \end{aligned}$$

Buscamos un mínimo en el intervalo  $-2 \leq x \leq 2$ .

Las primeras derivadas de la función son  $\nabla f(x) = 25x^4 + 6x^2 - 8x - 3$  y  $\nabla^2 f(x) = 100x^3 + 12x - 8$ .

Eligiendo un valor inicial de  $x_0 = 0$ , obtenemos:

$k$	0	1	2	3	4
$x_k$	-0.375	-0.2997	-0.29	-0.2899	-0.2899

Donde podemos ver que estamos convergiendo a la solución  $x = -0.2899$ .

Si lanzamos el método de Newton con un valor inicial distinto, de  $x_0 = 1$ , obtenemos:

$k$	0	1	2	3	4
$x_k$	0.8077	0.7141	0.6912	0.6899	0.6899

Ahora el método ha convergido a la solución  $x = 0.6899$ .

Buscamos un mínimo en el intervalo  $-2 \leq x \leq 2$ , por lo tanto deberíamos considerar el valor de la segunda derivada en los dos puntos críticos:

$$\nabla^2 f(-0.2899) = -13.9152 \quad \nabla^2 f(0.6899) = 33.115$$

Entonces podemos identificar que  $x = 0.6899$  es un mínimo local. ■

### 2.3.2 Método de gradiente

El método del gradiente es uno de los métodos más simples para resolver un problema de optimización no lineal sin restricciones, basado en el método de Newton. En comparación con el método de Newton, este método no requiere segundas derivadas, ni hacer la inversa de matrices. Por lo tanto, es más simple, y tiene menos coste computacional en cada iteración. La desventaja de este método es su convergencia, solo siendo lineal, lo cual hace que este método suele tardar en llegar a convergencia.

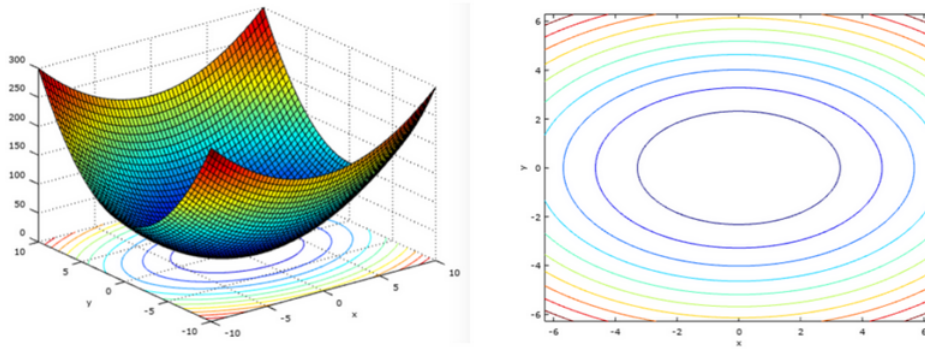
El método del gradiente toma un paso en cada iteración, en la dirección del gradiente de la función evaluada en el punto actual.

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

donde  $\alpha \in \mathbb{R}$ .

### ■ Ejemplo 2.7

Consideremos la función  $f(x, y) = x^2 + y^2$ .

Figura 2.2: Función  $f(x, y) = x^2 + y^2$ 

Buscamos el mínimo de esta función, usando el método del gradiente. Empezaremos desde el punto  $(x_0, y_0) = (4, 3)$ , y usaremos  $\alpha = 0.1$ . El gradiente de la función  $\nabla f(x, y) = (2x, 2y)$ .

Calculando los valores:

$$\begin{aligned} x_1 &= x_0 - 0.1 \nabla f(x_0) \\ x_1 &= (4, 3) - 0.1(2(4), 2(3)) \\ x_1 &= (3.2, 2.4) \\ x_2 &= x_1 - 0.1 \nabla f(x_1) \\ x_2 &= (3.2, 2.4) - 0.1(2(3.2), 2(2.4)) \\ x_2 &= (2.56, 1.92) \\ &\vdots \end{aligned}$$

Para el primer paso, vemos la dirección del primer paso por el gradiente de la función:

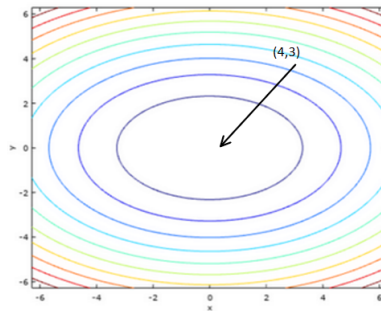


Figura 2.3: Dirección del gradiente

Si seguimos el método, llegaremos a la solución  $x^* = (0, 0)$ . Podemos comprobar con las técnicas de antes, que la función  $f(x, y)$  es convexa, y tiene un único óptimo en  $(0, 0)$ , lo cual es un mínimo.

■

La teoría de este método se puede resumir en lo siguiente:

#### Proposición 2.3.1

Sea  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  diferenciable en  $\hat{x}$ . Si existe  $d \in \mathbb{R}^n$  tal que  $\nabla f(\hat{x})^T d < 0$ , entonces existe  $\delta > 0$

tal que  $f(\hat{x} + \lambda d) < f(\hat{x})$  para  $\lambda \in (0, \delta)$ , es decir,  $d$  es una dirección de descenso de  $f$  en  $\hat{x}$ .

### Teorema 2.3.1

Sea  $f$  diferenciable en  $S$  abierto, no vacío tal que  $\nabla f(x) \neq 0$ . Entonces  $\bar{d} = -\frac{\nabla f(x)}{\|\nabla f(x)\|}$  es la dirección de máximo descenso en  $x$ .

**Nota** Dividimos por la norma del gradiente para no tener en cuenta el tamaño del gradiente, sino solo su dirección.

### Máximo Descenso

El método del gradiente puede ser mejorado, fijando en el tamaño del paso tomado en cada iteración del método,  $\alpha_k$ .

En cada iteración, el método del gradiente con máximo descenso busca un paso en la dirección que maximiza el descenso en la función objetivo.

### Algoritmo Máximo Descenso:

1. Sea  $\varepsilon > 0$  (escalar de parada). Elegir  $x_1, k = 1$ .
2. Si  $\|\nabla f(x_k)\| < \varepsilon$ , stop. Si no, sea  $d_k = -\nabla f(x_k)$ . Sea  $\lambda_k$  el óptimo de

$$\begin{aligned} &\text{Minimizar } f(x_k + \lambda d_k) \\ &\text{sujeto a } \lambda \geq 0 \end{aligned}$$

Sea  $x_{k+1} = x_k + \lambda_k d_k$ , hacer  $k = k + 1$  y repetir.

### ■ Ejemplo 2.8

Buscamos el valor de  $x$  que minimiza la ecuación cuadrática:

$$\begin{aligned} &\text{Minimizar } f(x) = \frac{1}{2}x^T Qx - c^T x \\ &\text{sujeto a } x \in \mathbb{R}^n \end{aligned}$$

con

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 25 \end{pmatrix}, \quad c = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}.$$

Aplicamos el método del gradiente. Buscando la dirección de descenso máxima:

$$d = -\nabla f = -(Qx - c)$$

Buscamos el paso de tomar en la dirección  $d$ , resolviendo el problema:

$$\begin{aligned} &\text{Minimizar } f(x + \lambda d) \\ &\text{sujeto a } \lambda \geq 0 \end{aligned}$$

Definimos  $g(\lambda) = f(x + \lambda d)$ , entonces:

$$g(\lambda) = \frac{1}{2}x^T Qx + \frac{1}{2}x^T Q\lambda d + \frac{1}{2}\lambda d^T Qx + \frac{1}{2}\lambda d^T Q\lambda d - c^T x - c^T \lambda d$$

Para hallar el mínimo de este problema, resolvemos  $\nabla g(\lambda) = 0$ :

$$\nabla g(\lambda) = \frac{1}{2}x^T Qd + \frac{1}{2}d^T Qx + d^T Q\lambda d - c^T x - c^T \lambda d$$

$$0 = x^T Qd + d^T Q\lambda d - c^T x - c^T \lambda d$$

$$\lambda = \frac{-d(x^T Q - c^T)}{d^T Qd}$$

$$\lambda = \frac{-d\nabla f(x)}{d^T Qd}$$

Para empezar el método, elegimos un punto inicial  $x_0 = (0, 0, 0)^T$ . Tenemos:

$$f(x_0) = 0, \quad \nabla f(x_0) = (1, 1, 1)^T, \quad \|\nabla f(x_0)\| = 1.7321.$$

Así calculamos el valor  $\lambda_0 = 0.0968$ , dando el valor nuevo  $x_1 = (-0.0968, -0.0968, -0.0968)^T$   
En paso  $k = 1$ , tenemos:

$$f(x_1) = -0.1452, \quad \nabla f(x_1) = (0.9032, 0.5161, -1.4194)^T, \quad \|\nabla f(x_1)\| = 1.7598.$$

Así calculamos el valor  $\lambda_1 = 0.059$ , dando el valor nuevo  $x_2 = (-0.15, -0.1272, -0.0131)^T$   
En paso  $k = 2$ , tenemos:

$$f(x_2) = -0.2365, \quad \nabla f(x_2) = (0.85, 0.3639, 0.6732)^T, \quad \|\nabla f(x_2)\| = 1.1437.$$

$$\vdots$$

(2.2)

El método tarda 216 iteraciones hasta conseguir un valor de  $\|\nabla f(x_k)\| < 10^{-8}$ . ■

**notas para considerar: punto inicial, no sea posible encontrar  $\lambda$  exacta, multimodal**

### ■ Ejemplo 2.9

Usa el método de gradiente para resolver el problema

$$\begin{aligned} &\text{Minimizar } f(x_1, x_2) = 4x_1^2 + 2x_2^2 + 4x_1x_2 - 3x_1 \\ &\text{sujeto a } \mathbf{x} \in \mathbb{R}^2 \end{aligned}$$

empezando desde el punto  $(2, 2)^T$  y realizando tres iteraciones.

La dirección de máximo descenso es  $d = (-8x_1 - 4x_2 + 3, -4x_1 - 4x_2)$ , y buscamos  $\lambda$  tal que

$$\begin{aligned} &\text{Minimizar } f(x + \lambda d) \\ &\text{sujeto a } \lambda \geq 0 \end{aligned}$$

donde  $g(\lambda) = f(x + \lambda d) = 4(x_1 + \lambda d_1)^2 + 2(x_2 + \lambda d_2)^2 + 4(x_1 + \lambda d_1)(x_2 + \lambda d_2) - 3(x_1 + \lambda d_1)$ .  
Derivando con respecto a  $\lambda$ , tenemos:

$$\nabla g(\lambda) = 8d_1(x_1 + \lambda d_1) + 4d_2(x_2 + \lambda d_2) + 4d_1x_2 + 4d_2x_1 + 8\lambda d_1d_2 - 3d_1$$

El mínimo valor de  $\lambda$  ocurre cuando  $\nabla g(\lambda) = 0$ . Entonces,

$$\lambda = \frac{3d_1 - x_1(8d_1 + 4d_2) - x_2(4d_2 + 4d_1)}{8d_1^2 + 4d_2^2 + 8d_1d_2}$$

Si empezamos con  $x^0 = (2, 2)^T$ , tenemos  $f(x^0) = 34$ ,  $\nabla f(x^0) = (21, 16)$ ,  $\|\nabla f(x^0)\| = 26.4$ ,  $d = (-21, -16)$  y  $\lambda = 0.09627$ .

La primera iteración con  $\lambda = 0.09627$  entonces nos da  $x^1 = x^0 + \lambda d = (-0.02167, 0.45968)$ .

En paso 2, tenemos  $\nabla f(x^1) = (-1.33464, 1.75204)$ ,  $\|\nabla f(x^1)\| = 2.2025$ ,  $d = (1.33464, -1.75204)$

y  $\lambda = 0.6202$ . Entonces  $x^2 = x^1 + \lambda d = (0.806074, -0.626935)$ .

En paso 3, tenemos  $\nabla f(x^2) = (0.940852, 0.716556)$ ,  $\|\nabla f(x^2)\| = 1.18265$ ,  $d = (-0.940852, -0.716556)$

y  $\lambda = 0.0963$ . Entonces  $x^3 = x^2 + \lambda d = (0.71547, -0.69594)$ .

Por ser un problema convexo y simple, podemos hallar su solución exacta, evaluando  $\nabla f(x) = 0$ .

Tenemos  $\nabla f(x) = (8x_1 + 4x_2 - 3, 4x_2 + 4x_1)$ , dando el sistema de ecuaciones:

$$8x_1 + 4x_2 - 3 = 0$$

$$4x_2 + 4x_1 = 0$$

con solución  $x = (0.75, -0.75)$ . Vemos que la solución del método del gradiente se está acercando a la solución exacta. ■

## 2.4 Problemas

1. Obténganse los extremos de las siguientes funciones

a.  $f(x_1, x_2) = 3x_1x_2 - x_1^3 - x_2^3$

b.  $f(x_1, x_2) = x_1x_2 + \frac{1}{x_1} + \frac{1}{x_2}$

c.  $f(x_1, x_2) = x_1^3 + x_2^3$

d.  $f(x_1, x_2) = 4x_1^2 + x_2^2 + 4x_1x_2$

e.  $f(x_1, x_2, x_3) = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}$

2. Justifíquese si

a. La función  $f(x_1, x_2) = x_2e^{x_1} + x_1x_2^2$  posee un mínimo en el punto  $(1, 1)$ .

b. La función  $f(x_1, x_2) = x_1x_2 - x_1^3x_2 - x_1x_2^3$  posee un máximo en el punto  $\frac{1}{2}, \frac{1}{2}$ .

3. Consideremos el ajuste de una recta por el método de los mínimos cuadrados. Dados los puntos  $(x_i, y_i) \in \mathbb{R}^2, i = 1, \dots, n$ , hállese una función  $f(x) = ax + b$   $a, b \in \mathbb{R}$  tal que la suma de las diferencias al cuadrado  $\sum_{i=1}^n (y_i - f(x_i))^2$  sea mínima.

4. Sea una empresa que produce tres bienes, cuyos precios de mercado son  $p_1 = 16, p_2 = 12$  y  $p_3 = 20$ . Su función de costes es

$$C(\mathbf{q}) = q_1^2 + 2q_2^2 + 3q_3^2 + 2q_1q_3 + 25,$$

donde  $q_1, q_2$  y  $q_3$  representan las cantidades producidas de cada uno de los bienes. Obténganse los valores de  $\mathbf{q}$  que maximizan el beneficio de la empresa.

5. Sea  $f(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4$ . ¿Cuál es la dirección de Newton en el punto  $x_0 = (0, 1)^T$ ?

6. Usando el método de Newton, resuelve el siguiente problema

$$\text{minimizar } f(x) = 5x^5 + 2x^3 - 4x^2 - 3x + 2$$

buscando la solución en el intervalo  $-2 \leq x \leq 2$ .

7. Usando el método de Newton, resuelve el siguiente problema

$$\text{minimizar } f(x_1, x_2) = 5x_1^4 + 6x_2^4 - 6x_1^2 + 2x_1x_2 + 5x_2^2 + 15x_1 - 7x_2 + 13.$$

usando el valor inicial  $(1, 1)^T$ . (Buscamos el mínimo, no el máximo)

8. Considere el problema

$$\text{minimizar } \frac{1}{2}x^T Qx - c^T x$$

donde  $Q$  es positiva definida, Demuestra que el método de Newton encontrará el mínimo en una iteración, desde cualquier punto inicial.

9. Use el método de gradiente para resolver

$$\text{minimizar } f(x_1, x_2) = 4x_1^2 + 2x_2^2 + 4x_1x_2 - 3x_1$$

con  $(x_1^0, x_2^0) = (2, 2)$ , usando 3 iteraciones.

10. Use el método del gradiente para resolver la ecuación cuadrática  $f(x) = \frac{1}{2}x^T Qx - c^T x$  donde

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & \gamma^2 \end{pmatrix}, \quad c = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

donde  $\gamma$  es un constante que puedes variar, por ejemplo  $\gamma = 1, 10, 100, 1000$  etc. (Use 4 iteraciones)

11. Consideremos el problema

$$\text{minimizar } f(x_1, x_2) = x_1^2 + 2x_2^2$$

a. Si el punto inicial es  $x_0 = (2, 1)^T$ , demuestra que la secuencia de puntos generado por el método del máximo descenso viene dada por

$$x_k = \begin{pmatrix} \frac{1}{3} \\ (-1)^k \end{pmatrix}$$

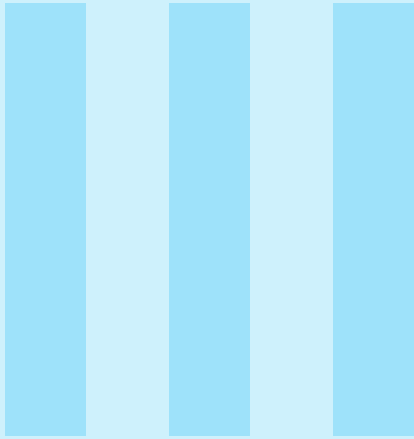
b. Demuestra que  $f(x_{k+1}) = \frac{f(x_k)}{9}$ .

12. Si aplicamos el método del gradiente para minimizar la función cuadrática

$$f(x) = \frac{1}{2}x^T Qx - c^T x$$

donde  $Q$  es positiva definida. Demuestra que:

$$\nabla f(x_{k+1}) = \nabla f(x_k) - \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T Q \nabla f(x_k)} Q \nabla f(x_k).$$



# Tema 3 - Optimización restringida

part.1chapter.1section.1.1section.1.2section.1.3subsection.1.3.1

- 3 Optimización restringida ..... 47**
- 3.1 Introducción
- 3.2 Restricciones de igualdad
- 3.3 Restricciones de desigualdad
- 3.4 Problemas

chapter.4section.4.1subsection.4.1.1subsection.4.1.2section\*.6se





## 3. Optimización restringida

### 3.1 Introducción

En este capítulo, pasamos a considerar problemas de optimización con restricciones de igualdad y desigualdad.

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{sujeto a:} & g_j(x) \leq 0 \quad j = 1, \dots, r \\ & h_i(x) = 0 \quad i = 1, \dots, m \\ & x \in X \end{array}$$

donde  $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ .

Diremos que  $x^*$  es un **punto regular** de (1) si  $\{\nabla h_i(x^*), i = 1, \dots, m, \nabla g_j(x^*), j \in I(x^*)\}$  son linealmente independientes.

#### ■ Ejemplo 3.1

Una fábrica produce cierto tipo de dispositivo que requiere acero como materia prima. Los costes sufridos en la fábrica son mano de obra, que cuesta 20 euros por hora y precio del acero, que cuesta 170 euro por tonelada. Si los ingresos de la fábrica,  $R$ , se modelan por la ecuación

$$R(h, s) = 200h^{2/3}s^{1/3}$$

donde  $h$  son las horas de trabajo y  $s$  son las toneladas de acero.

Si el presupuesto de la fábrica es 20,000 euros, ¿cuál es su máximo ingreso posible?

El problema para considerar entonces es:

$$\begin{array}{ll} \text{maximizar} & 200h^{2/3}s^{1/3} \\ \text{sujeto a:} & 20h + 170s = 20000 \end{array}$$

■

### 3.2 Restricciones de igualdad

En esta sección, concentramos en la definición y solución de problemas de optimización con restricciones de **igualdad**.

$$\begin{aligned} \text{Minimizar} \quad & f(x) \\ \text{sueto a :} \quad & h_i(x) = 0 \quad i = 1, \dots, m \end{aligned} \quad (3.1)$$

donde  $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ .

Ahora, el problema es encontrar máximos y mínimos de la función  $f$ , no en el conjunto  $S$  en el que está definida, sino en un subconjunto suyo donde los puntos de  $S$  satisfacen las restricciones  $h_i(x) = 0$ . Este subconjunto de  $S$  se le denomina el **conjunto factible**.

**Nota** Las soluciones de este tipo de problema serán, en general, distintas a las obtenidas del mismo problema sin restricciones.

#### ■ Ejemplo 3.2

Consideremos la función

$$f(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 1)^2,$$

que podemos visualizar en las siguientes gráficas.

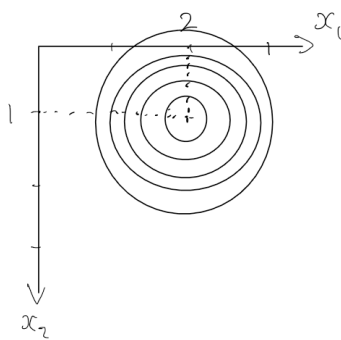


Figura 3.1: Curvas de nivel

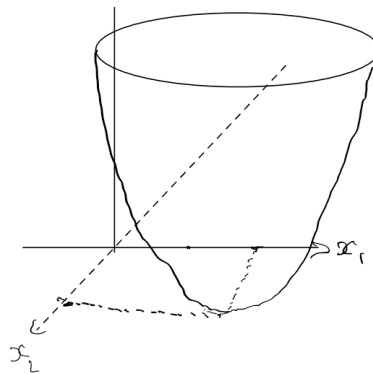


Figura 3.2: Función

Es claro que dicha función tiene un mínimo en el punto  $(x_1, x_2) = (2, 1)$ , con el valor óptimo  $f(x^*) = 0$ .

Añadimos la restricción  $x_1 + x_2 - 2 = 0$  al problema, que representamos gráficamente:

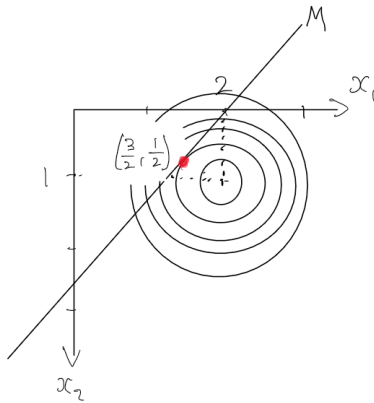


Figura 3.3: Curvas de nivel con restricciones

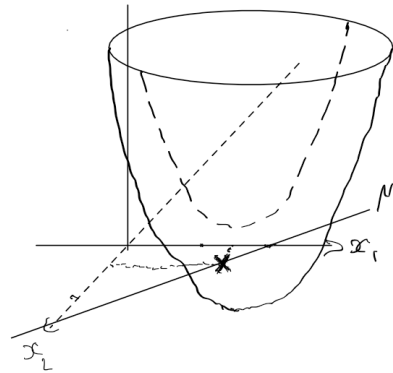


Figura 3.4: Función con restricciones

Ahora, el mínimo se encuentra en el punto  $(x_1, x_2) = (\frac{3}{2}, \frac{1}{2})$  con el valor óptimo  $f(x^*) = \frac{1}{2}$ .

■

La idea principal para la resolución de problemas de optimización con restricciones es de reducir el problema a uno equivalente, sin restricciones, que se le denomina el **problema reducido**.

### ■ Ejemplo 3.3

Consideremos el problema

$$\begin{aligned} \text{Minimizar} \quad & f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 \\ \text{sujeto a:} \quad & x_1 - x_2 + 2x_3 = 2 \end{aligned}$$

Usando la restricción en la forma  $x_1 = 2 + x_2 - 2x_3$ , podemos eliminar una variable del problema y quedar con un problema sin restricciones:

$$\text{Minimizar} \quad 2x_2^2 + 3x_3^2 - 4x_2x_3 + 2x_2.$$

Usando técnicas de resolver problemas sin restricciones de Tema 2, vemos que la solución de este problema es  $(x_2, x_3) = (-1.5, -1)$ , dando solución al problema original  $(x_1, x_2, x_3) = (2.5, -1.5, -1)$ .

■

## 3.2.1 Lagrange

Reducir un problema no es fácil, normalmente.

quizás necesito un ejemplo de un problema con alta dimensiones donde será difícil de conseguir esto.

Consideremos el problema

$$\begin{aligned} \text{Maximizar} \quad & f(x) = 48x + 96y - x^2 - 2xy - 9y^2 \\ \text{sujeto a:} \quad & 20x + 4y = 216, \\ & x, y \geq 0 \end{aligned}$$

La gráfica de las curvas de nivel de la función objetivo es:

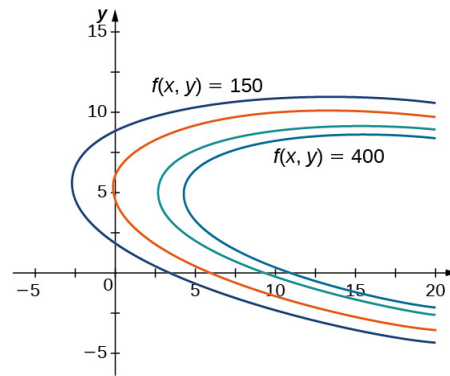


Figura 3.5: Curvas de nivel  $f(x, y)$  de  $c = 150, 250, 350, 400$ .

Por querer maximizar  $f$ , buscamos un valor lo más a la derecha posible. Hay que tener en cuenta la restricción, entonces dibujamos la línea  $20x + 40y = 216$ :

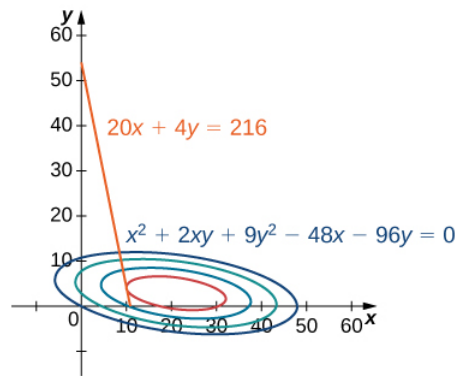


Figura 3.6: Curvas de nivel  $f(x, y)$  de  $c = 150, 250, 350, 400$ , con restricción.

Necesitamos que los valores de  $x, y$  están en la línea de esta restricción. Vemos que el valor máximo de la función objetivo ocurre cuando la línea de la restricción es tangente a una curva de nivel de la función.

Cuando dos vectores son tangentes (el gradiente es un vector en más de una variable), están apuntando en la misma dirección (o direcciones opuestas). Esto implica que uno es un múltiplo constante del otro (por apuntar en la misma dirección).

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$$

El principal resultado en el estudio de problemas de optimización con restricciones de igualdad es que el gradiente de la función objetivo reducido es una combinación lineal de los gradientes de las restricciones. Los coeficientes de esta combinación lineal se llaman *multiplicadores de Lagrange*. [en el libro p487-488, hay una gráfica y explicación](#)

### 3.2.2 Lagrange

Definimos la **función Lagrangiana** asociada a (3.1) como:

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x), \quad x \in X, \lambda \in \mathbb{R}^m$$

donde las restricciones  $h_i(x) = 0$ .

**Esto no es siempre cierto. Los puntos críticos de  $L$  son puntos de silla... también si el problema es convexo...**

De donde viene la lagrangiana? De la relación de los gradientes, con la condición necesaria de un punto crítico:  $\nabla L = 0 = \nabla f + \lambda \nabla h \Rightarrow \nabla f = -\lambda \nabla h$ . (ignoramos el signo, porque solo cambia el signo de  $\lambda$ ).

### 3.2.3 Condiciones necesarias/suficientes

**Teorema 3.2.1 — Condición necesaria de optimalidad. Teorema de los multiplicadores de Lagrange.**

Sea  $x^*$  un mínimo local y punto regular de (1). Entonces existe un único vector  $\lambda^* \in \mathbb{R}^m$ , denominado **multiplicador de Lagrange**, tal que

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \lambda^* \nabla h(x^*) = \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) = 0$$

Además si  $f$  y  $h$  son de clase  $C^2$ :

$$y^T (\nabla_{xx}^2 L(x^*, \lambda^*)) y = y^T \left( \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*) \right) y \geq 0 \quad \forall y : \nabla h(x^*)^T y = 0$$

**Nota** En estas definiciones, supongamos que  $h_i(x^*)$  son linealmente independientes (es decir,  $x^*$  es regular).

#### ■ Ejemplo 3.4

Consideremos el mismo problema del ejemplo anterior:

$$\begin{aligned} \text{Minimizar} \quad & f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 \\ \text{sujeto a:} \quad & x_1 - x_2 + 2x_3 = 2 \end{aligned}$$

Aplicamos el método de los multiplicadores de Lagrange. La función Lagrangiana es:

$$L(x, \lambda) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 + \lambda(x_1 - x_2 + 2x_3 - 2)$$

Calculamos el gradiente de la Lagrangiana

$$\nabla_x L = (2x_1 - 2 + \lambda, 2x_2 - \lambda, -2x_3 + 4 + 2\lambda)$$

Igualando a cero  $\nabla_x L = 0 \Rightarrow$

$$2x_1 - 2 + \lambda = 0, \quad 2x_2 - \lambda = 0, \quad -2x_3 + 4 + 2\lambda = 0$$

$$x_1 = 2 - \frac{1}{2}\lambda, \quad x_2 = \frac{1}{2}\lambda, \quad x_3 = 2 + \lambda$$

Sustituimos en la restricción  $x_1 - x_2 + 2x_3 = 2$  y vemos que  $\lambda = -3$ , dando punto crítico  $x^* = (2.5, -1.5, -1)$ .

Podemos comprobar la condición necesaria de segundo orden para clasificar el punto crítico. Si elegimos  $y^T = (0, 2, 1)$ , cumplimos  $\nabla h(x^*) y = 0$  donde  $\nabla h(x^*) = (1, -1, 2)$  y tenemos:

$$y^T \nabla_{xx}^2 L(x^*) y = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix} y = 6 \geq 0$$

Entonces, este punto es un mínimo local del problema. ■

**Nota** En el ejemplo anterior, la hessiana de la función es indefinida, aunque  $x^*$  es el mínimo. **ejemplo de ser necesaria pero no suficiente?**

**Teorema 3.2.2 — Condición suficiente de optimalidad.**

Supongamos que  $f$  y  $h$  son de clase  $C^2$  y sean  $x^* \in \mathbb{R}^n$  y  $\lambda^* \in \mathbb{R}^m$  tales que

$$\nabla L(x^*, \lambda^*) = 0$$

$$y^T \nabla_{xx}^2 L(x^*, \lambda^*) y > 0 \quad \forall y : \nabla h(x^*)^T y = 0$$

Entonces  $x^*$  es un mínimo local estricto para (1). De hecho existen unos escalares  $\sigma > 0$  y  $\varepsilon > 0$  tales que

$$f(x) \geq f(x^*) + \frac{\sigma}{2} \|x - x^*\|^2 \quad \forall x : h(x) = 0, \|x - x^*\| < \varepsilon$$

i.e., si la hessiana es definida positiva (valores propios positivos), tenemos un mínimo

■ **Ejemplo 3.5**

Consideremos el problema

$$\begin{aligned} \text{Minimizar} \quad & x^2 + y^2 \\ \text{sujeto a:} \quad & xy = 4 \end{aligned}$$

Sea  $f(x, y) = x^2 + y^2$  y  $h(x, y) = xy - 4$ , con  $\nabla f = (2x, 2y)$  y  $\nabla h = (y, x)$ . El Lagrangiano es

$$L(x, y; \lambda) = x^2 + y^2 + \lambda(xy - 4)$$

Igualando las derivadas a cero nos da el sistema

$$\begin{aligned} 2x + \lambda y &= 0 \\ 2y + \lambda x &= 0 \\ xy &= 4 \end{aligned}$$

con soluciones  $(2, 2; -2)$ ,  $(-2, -2; -2)$ . La hessiana de  $L$  es

$$H(L) = \begin{pmatrix} 2 & \lambda \\ \lambda & 2 \end{pmatrix}$$

Consideremos la condición de segundo orden  $\nabla h(x^*)^T y = 0$ , en ambas soluciones:

$$\nabla h(2, 2)^T y = 0 \quad \nabla h(-2, -2)^T y = 0$$

⇒

$$(2, 2)^T y = 0 \quad (-2, -2)^T y = 0$$

donde es fácil ver que ambas condiciones están satisfechas por los vectores  $(t, -t)$   $t \in \mathbb{R}$ . Entonces,

$$(t, -t) \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} (t, -t)^T = 8t^2 \quad (t, -t) \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix} (t, -t)^T = 0. \quad (3.2)$$

Vemos que el punto  $(2, 2)$  nos da la proyección de la hessiana semidefinida positiva, porque para cualquier valor de  $t$ ,  $8t^2$  sería positivo. Por lo tanto,  $(2, 2)$  es un mínimo. El punto  $(-2, -2)$  tiene valor 0 para cualquier valor de  $t$ , que podría ser un mínimo o máximo.

Si evaluamos la función objetivo en los dos puntos, vemos que ambos salen a 8, entonces ambos son mínimos locales. (Esto es una prueba con el valor de los puntos que se puede usar para distinguir entre máximos y mínimos. También veremos otro método con la **hessiana orlada**).

■

**Definición 3.2.1 — Hessiana Orlada.**

La **Hessiana Orlada** es una variante de la matriz hessiana normal, usada en problemas de optimización restringida. Los determinantes de sus menores principales se utilizan como criterio suficiente de segundo orden para clasificar puntos críticos.

$$\tilde{H}(L) = \begin{pmatrix} \frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda \partial x} \\ \frac{\partial^2 L}{\partial \lambda \partial x} & \frac{\partial^2 L}{\partial x^2} \end{pmatrix}$$

Consideremos los últimos  $n - k$  menores principales (donde  $n$  es el número de variables y  $k$  es el número de restricciones), y fijamos en sus signos:

- Si todos tienen signo  $(-1)^k$ , es un mínimo local.
- Si alternan en signo con el primero  $(-1)^{k+1}$ , es un máximo local.

**■ Ejemplo 3.6**

Siguiendo con el ejemplo anterior

$$\begin{array}{ll} \text{Minimizar} & x^2 + y^2 \\ \text{sujeto a:} & xy = 4 \end{array}$$

donde encontramos soluciones  $x_1^* = (2, 2; -2)$  y  $x_2^* = (-2, -2; -2)$ . Tenemos  $n = 2$  variables y  $k = 1$  restricciones. Por lo tanto, fijamos en el menor principal  $n - k = 1$ , es decir,  $H$ .

$$\tilde{H} = \begin{pmatrix} 0 & y & x \\ y & 2 & \lambda \\ x & \lambda & 2 \end{pmatrix}$$

Tenemos  $D_3 = -y(2y - \lambda x) + x(\lambda y - 2x)$ . Si metemos ambos puntos puntos críticos, sale  $D_3 = -32$ . Por tener signo negativo, cumple  $(-1)^k$ . Esto implica que son mínimos, comprobando el análisis anterior. ■

**■ Ejemplo 3.7**

Volvemos a mirar al ejemplo 3.4

$$\begin{array}{ll} \text{Minimizar} & f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 \\ \text{sujeto a:} & x_1 - x_2 + 2x_3 = 2 \end{array}$$

donde el punto crítico es  $x^* = (2.5, -1.5, -1)$  con  $\lambda = -3$ . Si calculamos la hessiana orlada:

$$\tilde{H} = \begin{pmatrix} 0 & 1 & -1 & 2 \\ 1 & 2 & 0 & 0 \\ -1 & 0 & 2 & 0 \\ 2 & 0 & 0 & -2 \end{pmatrix}$$

con  $n = 3$  y  $k = 1$ , tenemos que considerar  $n - k = 2$  menores principales,  $D_3$  y  $D_4$ . Vemos que  $D_3 = -4$  y  $D_4 = -8$ . Ambos menores principales son negativos, siguiendo la condición  $(-1)^k$ , entonces confirmamos que el punto crítico es un mínimo local. ■

Es bastante común en problemas de optimización que el método de los multiplicadores de Lagrange resulta en un sistema de ecuaciones difícil de resolver.

### ■ Ejemplo 3.8

Consideremos el problema:

$$\begin{aligned} \text{Minimizar} \quad & f(x) = x_1^4 x_2^2 + x_1^2 x_3^4 + \frac{1}{2} x_1^2 + x_1 x_2 + x_3 \\ \text{sujeto a:} \quad & x_1 + x_2 + x_3 = 1 \end{aligned}$$

Aplicamos el método de los multiplicadores de Lagrange. La función Lagrangiana es:

$$L(x, \lambda) = x_1^4 x_2^2 + x_1^2 x_3^4 + \frac{1}{2} x_1^2 + x_1 x_2 + x_3 + \lambda (x_1 + x_2 + x_3 - 1)$$

Calculamos el gradiente de la Lagrangiana

$$\nabla_x L = (4x_1^3 x_2^2 + 2x_1 x_3^4 + x_1 + x_2 + \lambda, 2x_1^4 x_2 + x_1 + \lambda, 4x_1^2 x_3^3 + 1 + \lambda)$$

Igualando a cero  $\nabla_x L = 0$  resulta en un sistema de ecuaciones no lineales difíciles de resolver. ■

### 3.2.4 Sensibilidad

En problemas de optimización, las ligaduras representan restricciones físicas que deseamos imponer en una situación. Es interesante considerar la sensibilidad del valor del óptimo  $x^*$  con respecto a cambios (pequeños) en las restricciones. Consideremos el problema general

$$\begin{aligned} \text{Minimizar} \quad & f(x) \\ \text{sujeto a:} \quad & Ax = b \end{aligned}$$

donde  $f$  es  $C^2$ , las restricciones son independientes y  $x^*$  es el valor óptimo con  $f(x^*)$  siendo el valor de la función en este punto.

Queremos saber el cambio en el valor de la función objetivo óptimo,  $f(x^*)$ , cuando ocurre un cambio en  $b$ . Es decir,  $\frac{df(x^*)}{db}$ . Si consideramos la función Lagrangiana en el punto óptimo, sabemos que tenemos que cumplir las restricciones, y que  $h_i(x^*) - b_i = 0 \quad \forall i$ . Por lo tanto,  $\frac{dL(x^*; \lambda^*)}{db} = \frac{df(x^*)}{db}$ .

Si ahora consideremos todo en función de  $b$ ,  $L(x^*(b); \lambda^*(b), b)$ , usamos la regla de la cadena para encontrar el cambio de  $L$  con respecto a  $b$ .

$$\frac{dL(x^*(b); \lambda^*(b), b)}{db} = \frac{\partial L}{\partial x} \frac{\partial x^*}{\partial b} + \frac{\partial L}{\partial \lambda} \frac{\partial \lambda^*}{\partial b} + \frac{\partial L}{\partial b} \frac{\partial b}{\partial b}$$

Por ser  $x^*$  óptimo,  $\nabla L = 0$ . Entonces podemos simplificar a

$$\frac{dL(x^*(b); \lambda^*(b), b)}{db} = \frac{\partial L}{\partial b}$$

Si la lagrangiana es  $L = f(x) + \lambda(h(x) - b)$ , vemos que  $\frac{\partial L}{\partial b} = -\lambda$ . Con  $\frac{dL(x^*; \lambda^*)}{db} = \frac{df(x^*)}{db}$ , llegamos a:

$$\frac{df(x^*)}{db} = -\lambda^*$$

También vemos que resulta  $df(x^*) = -db\lambda$ . El nuevo valor del óptimo se puede calcular con  $f_{\text{nuevo}}^* = f_{\text{antiguo}}^* + df(x^*)$ .

### ■ Ejemplo 3.9

Seguimos con el mismo ejemplo de antes:

$$\begin{aligned} \text{Minimizar} \quad & f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 \\ \text{sueto a :} \quad & x_1 - x_2 + 2x_3 = 2 \end{aligned}$$

que tiene como solución  $x^* = (2.5, -1.5, -1)$  con  $f(x^*) = -1.5$  y  $\lambda^* = -3$ . Consideremos el mismo problema, con un cambio en la restricción:

$$\begin{aligned} \text{Minimizar} \quad & f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 \\ \text{sueto a :} \quad & x_1 - x_2 + 2x_3 = 2 + \delta \end{aligned}$$

El mínimo de este problema es entonces  $f(\bar{x}) = -1.5 - 3\delta$ . ■

### ■ Ejemplo 3.10

Siguiendo con el primer ejemplo,

$$\begin{aligned} \text{maximizar} \quad & 200h^{2/3}s^{1/3} \\ \text{sueto a :} \quad & 20h + 170s = 20000 \end{aligned}$$

donde la solución es  $(h^*, s^*) = (\frac{2000}{3}, \frac{2000}{51})$ , con  $\lambda^* = -2.593$ . El ingreso total de este punto es  $R^* = 51,777$  euros.

Podemos plantear un cambio en el presupuesto de la fábrica. Si ahora pueden gastar hasta 25000 euros, ¿cuál sería su máximo ingreso?.

Vemos que  $\nabla b = 5000$ , y que  $R_{nuevo}^* = 51777 + 2.593 \cdot 5000 = 64742$  euros. ■

## 3.3 Restricciones de desigualdad

En esta sección, consideremos problemas de optimización que incluyen restricciones de desigualdad de la forma

$$\begin{aligned} \text{Minimizar} \quad & f(x) \\ \text{sueto a :} \quad & g_j(x) \leq 0 \quad j = 1, \dots, r \\ & h_i(x) = 0 \quad i = 1, \dots, m \\ & x \in X \end{aligned}$$

donde  $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ .

### ■ Ejemplo 3.11

$$\begin{aligned} \text{Minimizar} \quad & f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \\ \text{sueto a :} \quad & x_1 + 2x_2 \geq 2 \\ & x_1 - x_2 \geq -1 \\ & -x_1 \geq -3. \end{aligned}$$

El objetivo de este problema es encontrar el punto más cercano al origen de la función  $f$ . Si dibujamos la región factible:

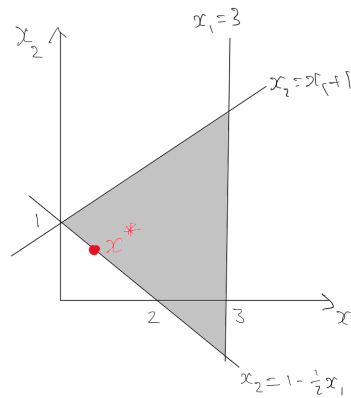


Figura 3.7: Restricciones de desigualdad.

Vemos que el mínimo de este problema va a estar por encima de la línea  $x_1 + 2x_2 = 2$ . Si suponemos que esto es cierto, podemos considerar el siguiente problema con restricciones de igualdad:

$$\begin{aligned} \text{Minimizar} \quad & f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \\ \text{sujeto a:} \quad & x_1 + 2x_2 = 2 \end{aligned}$$

Usando las condiciones de optimalidad de primer orden, vemos que la solución de este problema es  $x^* = (\frac{2}{5}, \frac{4}{5})^T$ ,  $f(x^*) = \frac{2}{5}$ ,  $\lambda^* = \frac{2}{5}$ . Dado que la función objetivo es convexa, es un mínimo global. Lo que queda por hacer para determinar que  $x^*$  es el mínimo del problema original es demostrar que el valor de la función no decrece cuando tomamos un paso pequeño hacia el interior de la región factible, donde  $x_1 + 2x_2 > 2$ . Si consideremos el cambio  $x_1 + 2x_2 = 2 + \delta$ , el cambio a la función sería  $\lambda^* \delta$ . Si el valor de  $\lambda^*$  es negativo, resultaría en un valor más pequeño de la función objetivo. Nosotros tenemos  $\lambda > 0$ , entonces podemos concluir que  $x^*$  es el mínimo del problema original. ■

En este ejemplo, hemos identificado las **restricciones activas** en la solución, y resuelto el problema con restricciones de igualdad. No siempre es fácil identificar el conjunto de **restricciones de desigualdad activas**

$$I(x) = \{j : g_j(x) = 0\}$$

y resulta necesario desarrollar condiciones más rigurosas para tratar de resolver estos problemas.

**Nota** Si  $j \notin I(x)$  diremos que  $g_j$  es una **restricción inactiva** en  $x$ .

### ■ Ejemplo 3.12

Seguimos con el mismo ejemplo de antes, pero suponemos que la tercera restricción es activa.

$$\begin{aligned} \text{Minimizar} \quad & f(x) = \frac{1}{2}x_1^2 + x_2^2 \\ \text{sujeto a:} \quad & -x_1 = -x_3 \end{aligned}$$

La solución de este problema es  $x_1 = 3, x_2 = 0, \lambda = -3$ . Entonces, por tener el multiplicador de Lagrange negativo, si tomamos un paso hacia el interior de la región factible, resultaría en un valor de la función objetivo más pequeño. Por lo tanto, no es el mínimo. ■

Lo que hemos definido es que si  $x^*$  es el óptimo del problema

$$\begin{aligned} &\text{Minimizar} && f(x) \\ &\text{sujeto a:} && g_j(x) \leq 0 \quad j = 1, \dots, r \\ &&& h_i(x) = 0 \quad i = 1, \dots, m \\ &&& x \in X \end{aligned}$$

y si el problema es regular y las restricciones inactivas se han eliminado,  $x^*$  también es el óptimo del problema

$$\begin{aligned} &\text{Minimizar} && f(x) \\ &\text{sujeto a:} && g_j(x) = 0 \quad \forall j \in I(x) \\ &&& h_i(x) = 0 \quad i = 1, \dots, m \\ &&& x \in X \end{aligned}$$

### 3.3.1 Karush-Kuhn-Tucker

El método demostrado en esta sección es una extensión del método de Lagrange para incluir restricciones de desigualdades.

$$\begin{aligned} &\text{Minimizar} && f(x) \\ &\text{sujeto a:} && g_j(x) \leq 0 \quad j = 1, \dots, r \\ &&& h_i(x) = 0 \quad i = 1, \dots, m \\ &&& x \in X \end{aligned}$$

Para seguir con la solución de este tipo de problema, hay que definir la forma **estándar**. Esto tiene que ver con el signo de las desigualdades en las restricciones, donde buscamos la forma:

$$g_j(x) \leq 0 \quad \forall j = 1, \dots, r$$

Definimos la función Lagrangiana

$$L(x; \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^r \mu_j g_j(x)$$

Consideramos las condiciones necesarias y suficientes para óptimos del problema de optimización, utilizando la función Lagrangiana. Lo siguiente es una extensión de las condiciones desarrolladas para problemas con restricciones de igualdad.

#### Condición Necesaria de Optimalidad

##### **Teorema 3.3.1 — Condición necesaria de Karush-Kuhn-Tucker.**

Sea  $x^*$  un mínimo local de (1) y supongamos que es regular. Entonces existen unos únicos vectores  $\lambda^* \in \mathbb{R}^m$  y  $\mu^* \in \mathbb{R}^r$  tales que

##### **Estacionaridad**

$$- \nabla_x L(x^*, \lambda^*, \mu^*) = 0,$$

##### **Factibilidad dual**

$$- \mu_j^* \geq 0 \quad j = 1, \dots, r,$$

##### **Factibilidad primal**

$$- h_i(x^*) = 0 \quad i = 1, \dots, m,$$

$$- g_j(x^*) \leq 0 \quad j = 1, \dots, r,$$

**Holgura complementaria**

$$- (\mu^*)^T g(x^*) = 0.$$

**Nota:** Si  $f(x)$  es convexa y  $g_j(x)$  son convexas, las condiciones necesarias son también suficientes.

la aplicación de Farkas para demostrar estacionaridad. factibilidad dual de antes en lineal. primal de las restricciones. Holgura complementaria - que las restricciones activas cumplen  $g = 0$ , mientras que las inactivas  $\lambda = 0$ . Uno de los dos es cero

**Definición 3.3.1 — Regularidad.**

Extendemos la definición de un punto **regular**. Un punto  $x^*$  es **regular** si para restricciones:

**de igualdad**  $\nabla h_i(x^*)$  son independientes linealmente,

**de desigualdad** los gradientes de las restricciones activas son independientes linealmente

$$\{\nabla g_j(x^*) : g_j(x^*) = 0 \forall j \in I\}.$$

**Demostración 3.1 — Explicación.**

El desarrollo de las condiciones necesarias y suficientes para la resolución de problema de optimización con restricciones de desigualdad depende de las condiciones anteriores, y también de los teoremas de separación. (hiperplanos y Farkas).

Empezamos con restricciones lineales, de desigualdad. Las condiciones para restricciones de igualdad son iguales que antes.

Buscamos la solución del problema

$$\begin{aligned} &\text{Minimizar} && f(x) \\ &\text{sujeto a :} && Ax \leq b \end{aligned}$$

donde  $A$  es una matriz  $r \times n$  con filas  $a_i^T$ . Suponemos que el sistema tiene una solución factible. Si  $x^*$  es una solución local, podemos decir que las restricciones inactivas no tienen efecto a la optimalidad de este punto, y las podemos descartar (por ahora). Sea  $\hat{A}$  la matriz de filas  $a_i^T$  de restricciones activas y  $\hat{b}$  el vector de valores de  $b$  correspondientes. Si  $x^*$  es una solución del problema original, entonces es una solución del problema

$$\begin{aligned} &\text{Minimizar} && f(x) \\ &\text{sujeto a :} && \hat{A}x = \hat{b} \end{aligned}$$

donde suponemos que las filas de  $\hat{A}$  son independientes (no necesario luego, pero simplifica la demostración).

Se puede llegar a cualquier otro punto factible de la región desde  $x^*$  por una dirección factible  $p$ . Una dirección es factible si  $\hat{A}p \geq 0$ .

**gráfica**

Las condiciones necesarias de problemas de igualdad nos dicen que  $\nabla f(x^*) = \hat{A}^T \hat{\lambda}^*$ , donde  $\hat{\mu}^*$  es el vector de multiplicadores de Lagrange de las restricciones activas. La condición necesaria

de segundo orden nos dice  $y^T \nabla_{xx}^2(x^*, \mu^*) y \geq 0 \quad \forall y : \nabla h(x^*)^T = 0$ .

Como hablado antes, un multiplicador de Lagrange negativo significaría una mejora en la función objetivo, con un paso hacia el interior de la región factible. Por lo tanto, necesitamos demostrar que siempre son positivos. Si  $\hat{\mu}^* \geq 0$  no es cierto, entonces existe al menos un elemento, por ejemplo el primero  $\hat{\mu}_1^*$ , negativo. Sea  $e_1$  un vector cuyo primer elemento es 1 y el resto ceros (tamaño  $1 \times$  número de restricciones activas).

Por ser las filas de  $\hat{A}$  independientes, podemos encontrar un vector  $p$  tal que  $\hat{A}p = e_1$ . Si  $\hat{A}p \geq 0$ ,  $p$  es una dirección factible. Pero,

$$p^T \nabla f(x^*) = p^T \hat{A}^T \hat{\mu}^* = e_1^T \hat{\mu}^* = \hat{\mu}_1^* < 0.$$

entonces  $p$  es una dirección factible de descenso en el punto  $x^*$ . Esto es una contradicción, por ser  $x^*$  un mínimo local por definición. Por lo tanto,  $\hat{\mu}^* \geq 0$ .

### ■ Ejemplo 3.13

Usamos las condiciones necesarias de KKT para buscar el máximo de  $f(x)$  en el siguiente problema:

$$\begin{aligned} \text{Maximizar} \quad & f(x) = \ln(x_1 + 1) + x_2 \\ \text{sujeto a:} \quad & 2x_1 + x_2 \leq 3 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Escribimos el problema en el formato de las condiciones necesarias de KKT:

$$\begin{aligned} \text{Maximizar} \quad & f(x) = \ln(x_1 + 1) + x_2 \\ \text{sujeto a:} \quad & 2x_1 + x_2 - 3 \leq 0 \\ & -x_1 \leq 0 \\ & -x_2 \leq 0 \end{aligned}$$

La función Lagrangiana:

$$L(x_1, x_2; \lambda) = \ln(x_1 + 1) + x_2 - \lambda_1(2x_1 + x_2 - 3) - \lambda_2(-x_1) - \lambda_3(-x_2)$$

Por la estacionaridad del problema, calculamos  $\nabla_x L = 0$ :

$$\frac{\partial L}{\partial x_1} = \frac{1}{x_1 + 1} - 2\lambda_1 + \lambda_2 = 0 \quad \frac{\partial L}{\partial x_2} = 1 - \lambda_1 + \lambda_3 = 0$$

y por la holgura complementaria:

$$\lambda_1(2x_1 + x_2 - 3) = 0 \quad \lambda_2(-x_1) = 0 \quad \lambda_3(-x_2) = 0$$

Intentamos resolver el sistema de ecuaciones. Notamos que si  $\lambda_1 = 0$ , entonces tendremos  $\lambda_3 = -1$ , lo cual no está permitido por factibilidad dual. Entonces, tenemos que  $2x_1 + x_2 - 3 = 0$ . Desde este punto, no es tan obvio lo que hay que hacer para avanzar.

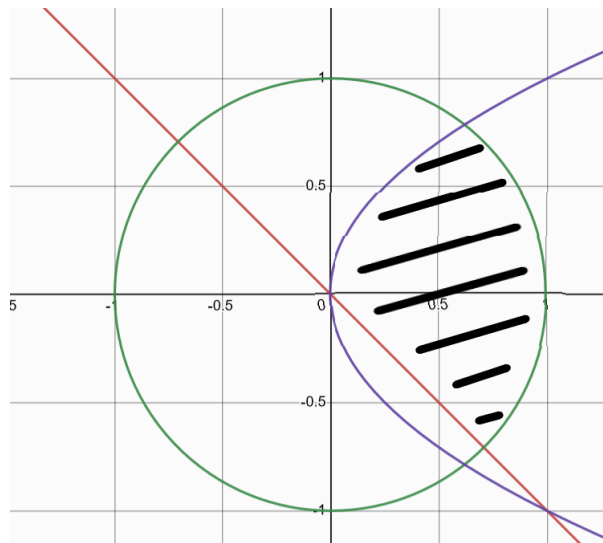
Si empezamos de nuevo y decir que  $\lambda_3 = 0$ , entonces tenemos que  $\lambda_1 = 1$ . Si  $x_1 = 0$ , vemos que  $\lambda_2 = 1$  y que  $x_2 = 3$ . Dando como solución  $(x_1, x_2) = (0, 3)$ . ■

falta explicar esto un poco más?

### ■ Ejemplo 3.14

Determinar el máximo y mínimo del siguiente problema de optimización. **di algo sobre la gráfica?**

$$\begin{aligned} \text{Maximizar} \quad & f(x) = x_2 \\ \text{sujeto a:} \quad & x_1^2 + x_2^2 \leq 1 \\ & -x_1 + x_2^2 \leq 0 \\ & x_1 + x_2 \geq 0 \end{aligned}$$



La función Lagrangiana:

$$L(x_1, x_2; \lambda_1, \lambda_2, \lambda_3) = x_2 + \lambda_1(x_1^2 + x_2^2 - 1) + \lambda_2(-x_1 + x_2^2) + \lambda_3(-x_1 - x_2)$$

Por la condición de estacionaridad del problema, calculamos  $\nabla_x = 0$ :

$$\frac{\partial L}{\partial x_1} = 2\lambda_1 x_1 - \lambda_2 - \lambda_3 = 0, \quad \frac{\partial L}{\partial x_2} = 1 + 2\lambda_1 x_2 + 2\lambda_2 x_2 - \lambda_3 = 0$$

y por la condición de la holgura complementaria:

$$\lambda_1(x_1^2 + x_2^2 - 1) = 0 \quad \lambda_2(-x_1 + x_2^2) = 0 \quad \lambda_3(-x_1 - x_2) = 0$$

Si  $\lambda_1 = 0$ ,

■

### 3.3.2 Condiciones Suficientes

#### Problemas convexos

En el caso de tener un problema convexo, las condiciones necesarias de KKT son también suficientes.

#### ■ Ejemplo 3.15

p113 ejemplo 3 del libro balbas gil

■

**Problemas no convexos**

Si el problema no es convexo, condiciones suficientes de segundo orden pueden ser usadas para confirmar la existencia de mínimos o máximos locales.

**Teorema 3.3.2 — Condición necesaria de segundo orden.**

Sea  $x^*$  una solución local (y punto regular) con  $\mu^*$  tal que las condiciones KKT están satisfechas y sea  $F(\mu^*)$  el conjunto:

$$w \in F(\mu^*) \iff \begin{cases} \nabla h_i(x^*)^T w = 0, & i \in M \\ \nabla g_i(x^*)^T w = 0, & i \in I(x^*), \mu^* > 0 \\ \nabla g_i(x^*)^T w \leq 0, & i \in I(x^*), \mu^* = 0 \end{cases}$$

Entonces  $w^T \nabla_{xx}^2 L(x^*, \mu^*) w \geq 0 \quad \forall w \in F(\mu^*)$ .

**Teorema 3.3.3 — Condición suficiente de segundo orden.**

Sea  $x^* \in \mathbb{R}^n$  un punto factible y  $\mu^*$  el multiplicador de Lagrange tal que las condiciones de KKT se satisfacen. Si  $(x^*, \mu^*)$  satisface la siguiente condición entonces  $x^*$  es un mínimo local estricto.

$$w^T \nabla_{xx}^2 L(x^*, \mu^*) w > 0 \quad \forall w \in F(\mu^*), w \neq 0.$$

**■ Ejemplo 3.16**

Considere el problema

$$\begin{aligned} \text{Min} \quad & f(x) = x^2 - y^2 \\ \text{sujeto a:} \quad & x - y = 1 \\ & x, y \geq 0 \end{aligned}$$

Halle los puntos críticos del problema con las condiciones KKT. Use las condiciones de segundo orden para verificar si alguno de los puntos corresponde a un mínimo local estricto.

Función Lagrangiana:  $L(x, y; \lambda, \mu_1, \mu_2) = x^2 - y^2 + \lambda(x - y - 1) - \mu_1 x - \mu_2 y$ .

Condiciones KKT:

- i) Estacionaridad:  $2x - \mu_1 + \lambda = 0, -2y - \mu_2 - \lambda = 0$
- ii) Factibilidad dual:  $\mu_1, \mu_2 \geq 0$
- iii) Holgura complementaria:  $\mu_1 x = 0, \mu_2 y = 0$

Dando punto KKT factible:  $(x^*, y^*, \lambda^*, \mu_1^*, \mu_2^*) = (1, 0, -2, 0, 2)$ . El problema es no convexo entonces consideramos las condiciones de segundo orden.

Tenemos  $(x^*, y^*) = (1, 0)$  entonces la restricción  $y \geq 0$  es activa y  $x \geq 0$  inactiva. Para la restricción activa  $\mu_2 = 2$  entonces comprobamos  $\nabla g_i(x^*)^T w = 0 \rightarrow -w_2 = 0$ .

Para la restricción de igualdad comprobamos  $\nabla h_i(x^*)^T w = 0 \rightarrow w_1 - w_2 = 0$ .

La única solución es  $w = (0, 0)$  por lo tanto no existe  $w \neq 0$  que no satisfice las condiciones  $\Rightarrow x^*$  es mínimo local estricto. (En este caso no hace falta evaluar la condición suficiente, porque no existen  $w$  distintos a cero.) ■

**Teorema 3.3.4**

Sean  $f, g$  y  $h$  sin ninguna hipótesis de diferenciabilidad y  $X$  un subconjunto cualquiera de  $\mathbb{R}^n$ . Sea  $x^*$  factible para (1),  $\lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^r, \mu_j^* \geq 0, (\mu^*)^T g(x^*) = 0$  tal que

$$L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*), \quad \forall x \in X, h(x) = 0, g(x) \leq 0,$$

entonces  $x^*$  es mínimo global de (1).

el uso de esto... en comparación con los límites de la función objetivo y también con los otros valores posibles en el dominio de la función...

### 3.4 Problemas

1. Consideremos el problema

$$\begin{aligned} \text{Minimizar} \quad & f(x) = x_1^2 + x_1^2 x_3^2 + 2x_1 x_2 + x_2^4 + 8x_2 \\ \text{sujeto a:} \quad & 2x_1 + 5x_2 + x_3 = 3 \end{aligned}$$

- Determine si los siguientes puntos son puntos estacionarios:  $(0, 0, 2)$ ,  $(0, 0, 3)$  y  $(1, 0, 1)$ .
- Determine que tipo de solución es cada punto estacionario.

2. Escribe el siguiente problema en forma estándar:

$$\begin{aligned} \text{Minimizar} \quad & f(x) = x_1^2 + x_1^2 x_3^2 + 2x_1 x_2 + x_2^4 + 8x_2 \\ \text{sujeto a:} \quad & 2x_1 + 5x_2 + x_3 = 3 \\ & x \geq 3 \end{aligned}$$

3. Use el método de los multiplicadores de Lagrange para hallar los óptimos de los siguientes problemas

- $f(x_1, x_2) = x_1 x_2^3$  sujeto a  $2x_1 + 3x_2 = 4$ ,
- $f(x_1, x_2) = x_1^2 + 2x_1 x_2 + x_2^2$  sujeto a  $3x_1^2 + x_2^2 = 9$ ,
- $f(x, y, z) = x^2 + y^2 + z^2$  sujeto a  $2x + y + 2z = 9$  y  $5x + 5y + 7z = 29$

4. Define los problemas reducidos de la pregunta anterior. ¿Su solución es la misma?

5. identifique las restricciones activas en el siguiente problema

$$\begin{aligned} \text{Max} \quad & x_1 \\ \text{sujeto a:} \quad & x_1 \geq 0 \\ & x_2 \leq 2 - x_1 \\ & \frac{x_1}{3} + x_2 \leq 1 \\ & x_2 \geq 0 \end{aligned}$$

6. Halla las dimensiones de la caja con más volumen si la area de su superficie es  $64\text{cm}^2$ .

7. Dada la función  $f(x, y) = x^4 + y^4 - 2x^2 - 2y^2 + 4axy + 3$ , determinar que tipo de punto crítico es el origen en función de los valores del parámetro  $a$ .

8. Determinar los valores de los parámetros  $a, b, c$  para que la función  $f(x, y) = ax^2 + by^2 - 2x - 2y + c$  tenga un mínimo local en  $(1, 2)$  y que verifique  $f(1, 2) = 15$ .

9. Busca los extremos de los siguientes problemas, utilizando la hessiana orlada.

- $f(x, y) = xy$  sujeto a  $h(x, y) = x + y - 6$ ,
- $f(x, y, z) = x^2 + y^2 + z^2$  sujeto a  $3x + y + z = 5$ ,  $x + y + z = 1$ ,
- $f(x, y) = x + y$  sujeto a  $x^2 + y^2 = 2$ .

10. Si  $x$  es el número de horas que un alumno dedica al ocio, y el número de horas que dedica a dormir siesta, y  $z$  el número de horas que estudia una asignatura, su nota final vendrá dada por la función  $f(x, y, z) = -x^2 - y^2 - z^2 + 3z - yz$ . Determine, si es posible, el número de horas que debe destinar a cada actividad para maximizar su nota, sabiendo que quiere dedicar a estudiar tanto tiempo como al ocio y descanso juntos. ¿Cuál es el valor de dicha nota máxima? Por último, ¿cuál es la variación aproximada de su nota si el alumno decide dedicar una hora más a estudiar respecto a la situación anterior? ¿Y si estudia dos horas más? Establece el número de horas que debe incrementar su estudio para obtener un 5 aproximadamente.
11. La temperatura en un punto  $(x, y)$  de una placa de metal es  $T(x, y) = 4x^2 - 4xy + 4y^2$ . Una hormiga camina sobre la placa alrededor de la circunferencia de radio 5 con centro en el origen. ¿Cuáles son las temperaturas máxima y mínima encontradas por la hormiga? ¿Cuál es la variación aproximada de la temperatura si se aumenta el radio en una unidad?
12. Una empresa quiere construir un radiotelescopio en un planeta recién descubierto. Para minimizar la interferencia, quiere colocarlo donde el campo magnético del planeta es más débil. El planeta es esférico, con un radio de 6 unidades. Con base en un sistema de coordenadas cuyo origen es el centro del planeta, la fuerza del campo magnético está dada por  $F(x, y, z) = 6x - y^2 + xz + 60$ . ¿Dónde se debe colocar el radiotelescopio? ¿Cuál es la fuerza del campo magnético en tal caso? ¿Cuál es la variación aproximada de dicho campo magnético si el radio del planeta fuera 9?
13. Escribe las condiciones de KKT para el siguiente problema:

$$\begin{aligned} \text{Min } & c^T x \\ \text{s.a. } & Ax \leq b \\ & x \geq 0 \end{aligned}$$

14. Resuelve el problema

$$\begin{aligned} \text{Min } & -x_1 - 3x_2 \\ \text{s.a. } & x_1 - 2x_2 \geq -4 \\ & -x_1 - x_2 \geq -4 \\ & x_1, x_2 \geq 0 \end{aligned}$$

# IV

## Tema 4 - Dualidad Lagrangiana

part.1chapter.1section.1.1section.1.2section.1.3subsection.1.3.1

<b>4</b>	<b>Dualidad</b> .....	<b>67</b>
4.1	Introducción	
4.2	Problemas	
	chapter.5section.5.1section.5.2section.5.3section.5.4	





## 4. Dualidad

### 4.1 Introducción

Consideremos problemas de optimización de la forma

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{sujeto a:} & g_j(x) \leq 0 \quad j = 1, \dots, r \\ & h_i(x) = 0 \quad i = 1, \dots, m \\ & x \in X \end{array}$$

donde  $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ .

La condición necesaria de optimalidad de Karush-Kuhn-Tucker nos da, desde la condición de estacionaridad

$$\nabla_x L(x^*, \lambda^*, \mu^*) = 0$$

una función que depende del vector óptimo  $x^*$  y los vectores de multiplicadores de lagrange  $\lambda^*$  y  $\mu^*$ . Si tuviéramos los valores de  $\lambda^*$  y  $\mu^*$  a priori, sería más fácil (normalmente) resolver el problema restringido. En este capítulo, tratamos de definir un problema de optimización asociado al original, pero en vez de optimizar por  $x$ , optimizamos por  $\lambda$  y  $\mu$ .

#### Definición 4.1.1

A los multiplicadores de lagrange se le llamen las **variables dual**.

Estudiamos las condiciones para que las soluciones de ambos problemas sean iguales.

El óptimo de nuestro problema original ocurre en el vector  $x^*$ . Si pensamos en la factibilidad de sistemas, utilizando Farkas, podemos representar nuestro problema por sistemas parecidos:

$$f(x) \leq c, \quad g_i(x) \leq 0, \quad i = 1, \dots, m$$

y también

$$f(x) < c, \quad g_i(x) \leq 0, \quad i = 1, \dots, m$$

donde  $c \in \mathbb{R}$ .

Si  $x^*$  es óptimo, el primer sistema tiene solución en  $c = f(x^*)$ , y el segundo no tiene solución ( $x^*$  mínimo, entonces  $f(x) < c$  no existe porque  $c = f(x^*)$  es el mínimo). Este resultado es la base para la **dualidad lineal**.

Podemos considerar, **sin** linealidad, el sistema

$$(I) \quad \begin{aligned} f(x) &< c \\ g_i(x) &\leq 0 \quad i = 1, \dots, m \end{aligned}$$

donde  $f, g \in \mathbb{R}$ . No asumimos convexidad.

Si consideramos la función Lagrangiana: si existen  $\mu_i \geq 0$  tales que

$$f(x) + \sum \mu_i g_i(x) < c \quad (1)$$

no tiene soluciones, entonces (I) no tiene solución.

(porque  $\sum \mu g$  es negativo, si no tiene solución  $f(x)$  tampoco tiene)

Cuando tiene 1) solución? Cuando:

$$f(x) + \sum \mu_i g_i(x) \geq c$$

Entonces, tenemos una condición necesaria y suficiente. Hemos identificado una ecuación, la **Lagrangiana**, mayor o igual que el mínimo. Preguntamos, ¿cuál sería el *mejor* valor de  $x^*$  para conseguir el valor mínimo? el ínfimo.

Diferencia entre mínimo y ínfimo? El mínimo se obtiene, el ínfimo no hay por que obtener.

**Proposition 4.1.1 — Condición suficiente para factibilidad de (I).**

Considerando el sistema (I), y un sistema

$$(II) \quad \begin{aligned} \inf_{x \in X} \{f(x) + \sum_{i=1}^m \mu_i g_i(x)\} &\geq c \\ \mu_i &\geq 0 \quad i = 1, \dots, m \end{aligned}$$

con solución. Entonces (I) no tiene solución.

(porque si  $\mu g$  es negativo,  $f(x) > c$ ).

El resultado anterior nos lleva a una aplicación de la función Lagrangiana de nuestro problema.

**Definición 4.1.2 — Función Dual.**

Definimos la **función dual** de nuestro problema primal como

$$L^*(\lambda, \mu) = \inf_x \{f(x) + \sum \lambda h(x) + \sum \mu g(x)\} \quad (4.1)$$

donde  $\mu \geq 0$ .

La función dual nos da una cota inferior del valor óptimo del problema primal. i.e.  $L^*(\lambda, \mu) \leq c^*$ .

**Demostración 4.1** Sea  $\bar{x}$  un punto factible del problema primal. i.e.  $g_i(\bar{x}) \leq 0$  y  $h_j(\bar{x}) = 0$  con  $\mu \geq 0$ . Entonces  $\sum_{i=1}^m \mu g_i(\bar{x}) + \sum_{j=1}^k \lambda h_j(\bar{x}) \leq 0$ .

Entonces,

$$L(\bar{x}; \lambda, \mu) = f(\bar{x}) + \sum_{i=1}^m \mu g_i(\bar{x}) + \sum_{j=1}^k \lambda h_j(\bar{x}) \leq f(\bar{x})$$

$\Rightarrow$

$$L^*(\lambda, \mu) = \inf_x \{L(x; \lambda, \mu)\} \leq L(\bar{x}; \lambda, \mu) \leq f(\bar{x})$$

Si esto se cumple por cualquier punto factible  $\bar{x}$ , tenemos:

$$L^*(\lambda, \mu) \leq c^*.$$

Entonces tenemos una cota inferior del problema primal. ¿Cuál será la mejor cota inferior? Esta pregunta es igual a resolver el **problema dual**.

**Definición 4.1.3 — Problema Dual.**

El **problema dual** es el problema de optimización

$$\begin{array}{ll} \text{Maximizar} & L^*(\lambda, \mu) \\ \text{s.a.} & \mu \geq 0 \end{array}$$

El valor óptimo del problema dual también sería una cota inferior del problema primal. (o igual, como veremos más tarde)

**demostración?**

#### ■ Ejemplo 4.1

Sacar el problema dual del siguiente problema primal;

$$\begin{array}{ll} \text{Minimizar} & 2x_1^2 + x_2^2 \\ \text{s.a.} & x_1 + x_2 \leq 4 \end{array}$$

Función Lagrangiana:

$$L(x; \mu) = 2x_1^2 + x_2^2 + \mu(x_1 + x_2 - 4)$$

Aplicando la primera condición necesaria:

$$\nabla L(x; \mu) = (4x_1 + \mu, 2x_2 + \mu) = 0 \quad \Rightarrow \quad x_1 = \frac{-\mu}{4}, x_2 = \frac{-\mu}{2}$$

Función dual:

$$L^*(\mu) = 2\left(\frac{-\mu}{4}\right)^2 + \left(\frac{-\mu}{2}\right)^2 + \mu\left(\frac{-\mu}{4} + \frac{-\mu}{2} - 4\right)$$

Problema dual:

$$\begin{array}{ll} \text{Maximizar}_{\mu} & L(x; \mu) \\ \text{s.a.} & \mu \geq 0 \forall i \end{array}$$

**Proposition 4.1.2**

La función Lagrangiana  $L$  es una función lineal en  $\lambda$  y  $\mu$ , entonces cóncava y convexa a la vez. La función dual  $L^*$  es el mínimo de las funciones lineales, entonces siempre es cóncava. (no importa la conavidad de  $f(x)$ )

**Demostración 4.2**

Si por  $y \in \Omega$  tenemos  $f(x, y)$  cóncava en  $x$ , sea

$$g(x) = \inf_{y \in \Omega} \{f(x, y)\}$$

entonces el dominio de  $g(x)$  es cóncava, por ser intersección de cóncavas (lineales en  $\lambda, \mu$ ). Entonces si  $t \in [0, 1]$  y  $x_1, x_2 \in \text{dom}(g)$ :

$$g(tx_1 + (1-t)x_2, y) = \inf_{y \in \Omega} \{f(tx_1 + (1-t)x_2, y)\}$$

por cóncavidad *def*

$$g(tx_1 + (1-t)x_2, y) \geq \inf_{y \in \Omega} \{tf(x_1, y) + (1-t)f(x_2, y)\}$$

$$g(tx_1 + (1-t)x_2, y) \geq \inf_{y \in \Omega} \{tf(x_1, y)\} + \inf_{y \in \Omega} \{(1-t)f(x_2, y)\}$$

$$g(tx_1 + (1-t)x_2, y) \geq tg(x_1) + (1-t)g(x_2).$$

entonces  $g(x)$  es cóncava también. ■

**■ Ejemplo 4.2**

Consideremos el problema

$$\begin{array}{ll} \min & f(x) = x^2 \\ \text{s.a.} & x \geq 1 \end{array}$$

convertimos a la forma estándar

$$\begin{array}{ll} \min & f(x) = x^2 \\ \text{s.a.} & 1 - x \leq 0 \end{array}$$

que tiene solución  $x^* = 1, \mu^* = 2$  por el método de Lagrange.

La función Lagrange es  $L(x, \mu) = x^2 + \mu(1 - x)$  con  $\mu \geq 0$ , dando función dual:

$$L^*(\mu) = \inf_x \{x^2 + \mu(1 - x)\}$$

El ínfimo en el problema dual es un problema de minimización de una función de una variable real (positiva). Si calculamos la derivada  $\nabla L(x) = 2x - \mu$ , y aplicamos la primera condición necesaria de optimalidad  $\nabla L(x) = 0$ , tenemos  $x = \frac{\mu}{2}$ .

Entonces, la función dual sería  $L^*(\mu) = \mu - \frac{1}{4}\mu^2$ . Dando problema dual

$$\begin{array}{ll} \max & L^*(\mu) = \mu - \frac{1}{4}\mu^2 \\ \text{s.a.} & \mu \geq 0 \end{array}$$

Intentamos resolver este problema, aplicando la primera condición necesaria de optimalidad.  $\nabla L^*(\mu) = 1 - \frac{1}{2}\mu \Rightarrow \mu = 2$ . Entonces, el valor de la función dual es  $L^*(2) = 1$ , lo cual es igual al valor del óptimo del problema primal. ■

### ■ Ejemplo 4.3

También es posible que no siempre se puede formular el problema dual explícitamente.

Consideremos el problema

$$\begin{aligned} \min \quad & f(x) = e^x \\ \text{s.a.} \quad & 1 - x^2 \geq 0 \end{aligned}$$

con función dual  $L^*(\mu) = \inf_x \{e^x + \mu(x^2 - 1)\}$ , donde  $\mu \geq 0$ . El ínfimo en  $x$  requiere calcular la derivada  $\nabla L(x) = e^x + 2\mu x$ , donde vemos que no hay forma de despejar  $x$  en términos de  $\mu$ . Si formamos el problema dual de todas formas:

$$\begin{aligned} \max \quad & L^*(\mu) = e^x + \mu(x^2 - 1) \\ \text{s.a.} \quad & \mu \geq 0 \\ & e^x + 2\mu x = 0 \end{aligned}$$

podríamos intentar resolver este problema por el método de Lagrange.

Si en vez de esto, volvemos al problema primal y cambiamos la forma de la restricción  $1 - x^2 \geq 0$  a  $-1 \leq x \leq 1$ , tenemos:

$$\begin{aligned} \min \quad & f(x) = e^x \\ \text{s.a.} \quad & x - 1 \leq 0 \\ & -1 - x \leq 0 \end{aligned}$$

con función dual  $L^* = \inf_x \{e^x + \mu_1(x - 1) + \mu_2(-1 - x)\}$ . Buscamos el ínfimo de la función lagrangiana con la primera condición necesaria de optimalidad  $\nabla L = e^x + \mu_1 - \mu_2$

*Rightarrow*  $x = \ln(\mu_2 - \mu_1)$ . Entonces tenemos el problema dual

$$\begin{aligned} \max \quad & L^*(\mu) = \mu_2 - \mu_1 + \mu_1(\ln(\mu_2 - \mu_1) - 1) + \mu_2(-1 - \ln(\mu_2 - \mu_1)) \\ \text{s.a.} \quad & \mu \geq 0 \end{aligned}$$

donde ha sido posible despejar  $x$  y conseguir el problema dual bien definido. Podríamos seguir con la resolución desde este punto. ■

La solución del problema dual es una cota inferior del problema primal, como hablado antes. Consideremos la teoría sobre cuando son iguales.

#### **Teorema 4.1.1 — Dualidad débil.**

Sea  $x$  una solución factible al problema primal, y  $(\bar{x}; \bar{\lambda}, \bar{\mu})$  una solución factible al problema dual. Entonces,

$$f(\bar{x}) + \bar{\lambda}h(\bar{x}) + \bar{\mu}g(\bar{x}) \leq f(x).$$

La diferencia entre la solución del problema primal y dual siempre es positivo  $p^* - d^* \geq 0$ , y se le llama **holgura dual**.

### ■ Ejemplo 4.4

$$\begin{aligned} \min \quad & f(x) = -x^2 \\ \text{s.a.} \quad & x = 1 \\ & x \in X \end{aligned}$$

donde  $X = \{x : 0 \leq x \leq 2\}$ . La solución (por lagrange) es  $x^* = 1$  y  $f(x^*) = -1$ .

La lagrangiana es  $L(x; \lambda) = -x^* + \lambda(x - 1)$ , con gradiente  $\nabla L = -2x + \lambda$ .  $\nabla = 0 \Rightarrow x = \frac{\lambda}{2}$ .

Entonces la función dual es  $L^*(\lambda) = \lambda(\frac{\lambda}{4} - \lambda)$ . Si resolvemos el problema dual,  $\nabla L^*(\lambda) = 0 \Rightarrow \lambda = 2$ . Si calculamos  $L^*(2) = -2$ , vemos que tenemos holgura dual de 1. ■

#### ■ Ejemplo 4.5

■

#### Proposition 4.1.3

En problemas convexos, es decir problemas con función objetivo convexa y restricciones convexas, la holgura dual puede ser igual a cero, bajo ciertas condiciones. Una de ellas es la **condición de Slater**.

#### Definición 4.1.4 — Condición suficiente de Slater.

Sea el problema convexo

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & g_j(x) \leq 0 \quad j = 1, \dots, m \\ & Ax = b, Dx \leq d \\ & x \in X \end{aligned}$$

donde  $X \subset \mathbb{R}^n$  y  $g_1, \dots, g_m$  funciones convexas no-lineales definidas en  $X$ . El problema cumple la condición de Slater si  $\exists x \in X$  tal que  $g_i < 0 \forall i$ .

#### ■ Ejemplo 4.6

El problema

$$\begin{aligned} \min \quad & x_1 + e^{x_2} \\ \text{s.a} \quad & 3x_1 - 2e^{x_2} \geq 10 \\ & x_2 \geq 0 \end{aligned}$$

cumple la condición de Slater. Por ejemplo  $x = (5, 0)$ . ■

#### Teorema 4.1.2 — Dualidad fuerte.

Si el problema primal es convexo, y cumple la **condición de Slater**, tenemos que  $d^* = p^*$ , es decir, la holgura dual es cero. Entonces, las soluciones de los problemas primal y dual son iguales.

#### ■ Ejemplo 4.7

Consideremos el problema

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.a} \quad & -x_1 - x_2 + 4 \leq 0 \end{aligned}$$

donde  $x_1, x_2 \geq 0$ .

La solución (por KKT) es  $x^* = (2, 2)$  con  $f(x^*) = 8$ . Vemos que el problema cumple la condición de Slater, por ejemplo  $x = (3, 3)$ . Entonces, la solución del problema dual es igual a la solución primal.

La lagrangiana es  $L(x; \mu) = x_1^2 + x_2^2 + \mu(-x_1 - x_2 + 4)$ , con gradiente  $\nabla_x L(x; \mu) = (2x_1 - \mu, 2x_2 - \mu)$ . Con la primera condición de optimalidad, vemos que el ínfimo nos da  $x_1 = x_2 = \frac{\mu}{2}$ . Entonces, la función dual es  $L^*(\mu) = \frac{\mu^2}{2} + \mu(4 - \mu)$ .

Resolviendo el problema dual de  $\max L^*(\mu)$ , la primera condición de optimalidad dice  $\nabla L^*(\mu) = 0$ , dando  $0 = \mu + 4 - 2\mu \Rightarrow \mu = 4$ . El valor de la función dual en este punto es  $L^*(\mu) = 8$ . ■

#### 4.1.1 No solución

Puede ser que el problema primal tenga solución, y el dual no:

##### ■ Ejemplo 4.8

Resolvemos el problema primal

$$\begin{aligned} \min \quad & x + 2010 \\ \text{s.a} \quad & \frac{1}{2}x^2 \leq 0 \end{aligned}$$

Formamos la función lagrangiana  $L(x; \mu) = x + 2010 + \mu(\frac{1}{2}x^2)$ , con  $\nabla_x L(x; \mu) = 1 + \mu x$ . Desde la primera condición necesaria de optimalidad, vemos que  $0 = 1 + \mu x$ . Desde la holgura complementaria,  $\mu(\frac{1}{2}x^2) = 0$ . Si  $\mu = 0$ , llegamos a  $1 = 0$ , entonces sabemos que  $\frac{1}{2}x^2 = 0$ . Dando como solución única  $x^* = 0$ .

Formamos la función dual  $L^* = \inf_x \{L(x; \mu)\} = \frac{-1}{\mu}$ . El problema dual es

$$\begin{aligned} \max \quad & \frac{-1}{\mu} + 2010 + \frac{1}{2\mu} \\ \text{s.a} \quad & \mu > 0 \end{aligned}$$

La primera condición necesaria nos da  $\nabla_\mu L^*(\mu) = \frac{1}{2\mu^2}$ . Usando la primera condición necesaria de optimalidad  $\nabla_\mu L^* = 0$  nos lleva a  $0 = \frac{1}{2\mu^2}$ , lo cual no tiene solución. ■

Puede ser que el problema primal no tenga solución, y el dual sí:

##### ■ Ejemplo 4.9

Resolvemos el problema primal

$$\begin{aligned} \min \quad & e^{-x} \\ \text{s.a} \quad & x \geq 0 \end{aligned}$$

donde la función lagrangiana es  $L(x; \mu) = e^{-x} + \mu(-x)$ . La primera condición necesaria de optimalidad  $\nabla_x L(x; \mu) = -e^{-x} - \mu = 0$ . Llegamos a  $-x = \ln(-\mu)$ , donde tenemos  $\mu \geq 0$ , pero  $\nexists \ln(x)$  con  $x \leq 0$ . Por lo tanto, no hay solución.

Si formamos la función dual  $L^*(\mu) = \inf_x \{e^{-x} - \mu x\}$ . Desde  $\nabla_x L(x; \mu)$ , llegamos a  $x = -\ln(-\mu)$ , entonces  $L^*(\mu) = -\mu + \mu \ln(-\mu)$ . El problema dual es entonces

$$\begin{aligned} \max \quad & -\mu + \mu \ln(-\mu) \\ \text{s.a} \quad & \mu > 0 \end{aligned}$$

Tenemos  $\nabla_\mu L^*(\mu) = -1 + \ln(-\mu) - 1$ , y con la condición necesaria de optimalidad, tenemos  $2 = \ln(-\mu)$ . Entonces,  $\mu = -e^2$ , y el valor de  $L^*(-e^2) = 3e^2$ . ■

### 4.1.2 Explicación Geométrica

Resumen del problema primal y dual:

Primal

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & g_i(x) \leq 0 \quad i = 1, \dots, m \\ & h_j(x) = 0 \quad j = 1, \dots, k \\ & x \in X \subseteq \mathbb{R}^n \end{aligned}$$

Dual

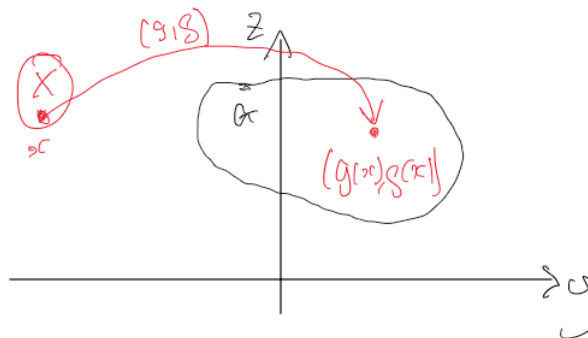
$$\begin{aligned} \max \quad & L^*(\lambda, \mu) \\ \text{s.a} \quad & \mu \geq 0 \end{aligned}$$

donde  $L^*(\lambda, \mu) = \inf_x \{f(x) + \sum \mu g(x) + \sum \lambda h(x) : x \in X\}$  es la función dual.

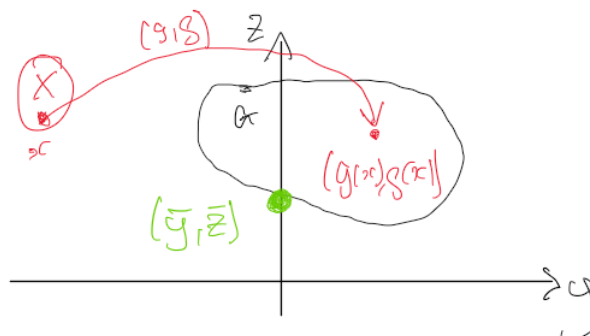
Un ejemplo geométrico: si el primal es:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & g(x) \leq 0 \\ & x \in X \end{aligned}$$

donde  $f : \mathbb{R}^n \rightarrow \mathbb{R}, g : \mathbb{R}^n \rightarrow \mathbb{R}$ . Definimos el conjunto  $G = \{(y, z) : y = g(x), z = f(x), x \in X\}$ . Es decir,  $G$  es el imagen de  $X$ , bajo el mapa  $(g, f)$ .



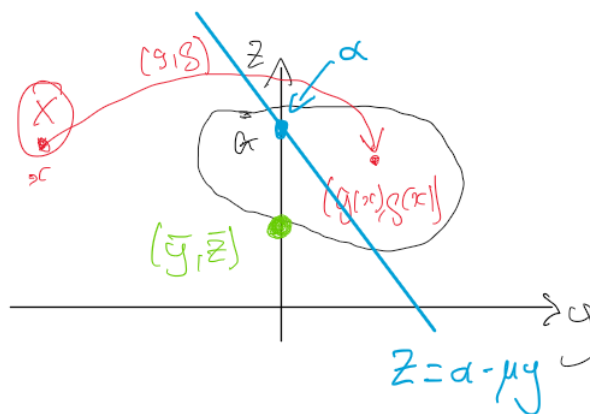
Entonces, el problema primal consiste en encontrar  $y \leq 0$  que tiene el mínimo de  $z$ :



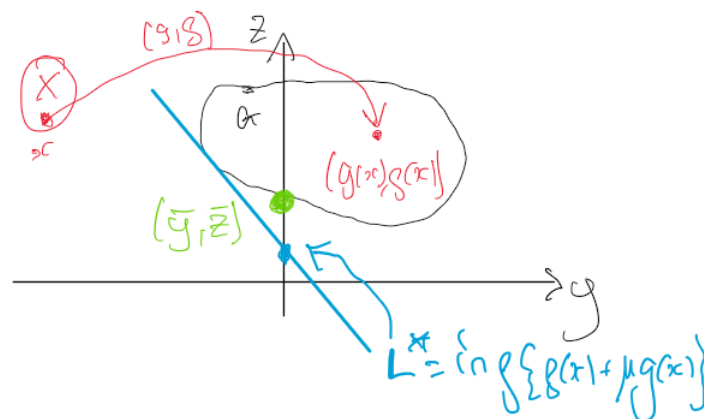
este punto es  $(\bar{y}, \bar{z})$ .

En el problema dual, con  $\mu \geq 0$ , la función dual es equivalente a minimizar  $z + \mu y$  en  $G$ . ( $\inf\{f(x) + \mu g(x)\} = \{z + \mu y\}$ ).

$z + \mu y$  es una ecuación lineal, que podemos considerar con  $z + \mu y = \alpha$ , donde  $\alpha$  es su corte en el eje de  $z$ , y tiene pendiente  $-\mu$ :  $z = \alpha - \mu y$

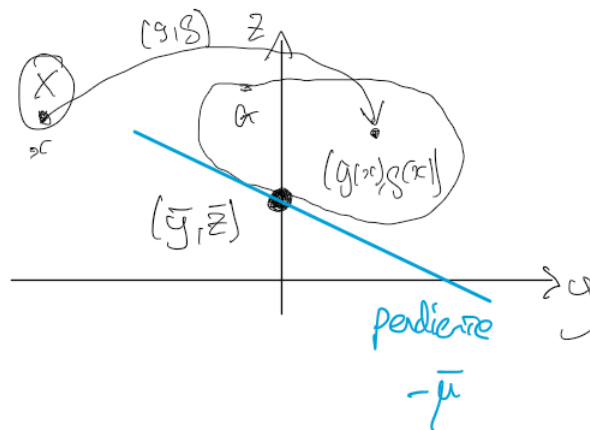


Para minimizar esta recta en  $G$ , tenemos que moverla hacia abajo, manteniendo contacto con  $G$ .



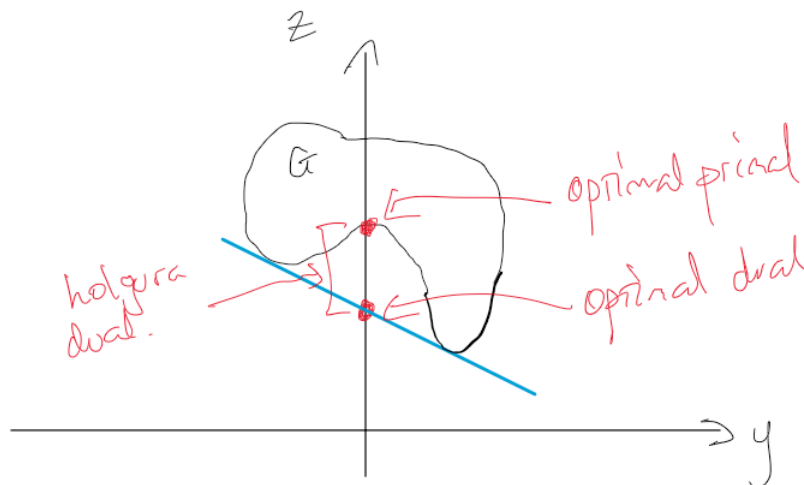
El corte con el eje de  $z$  es el valor de  $L^*$  cuando  $\mu \geq 0$ .

Para resolver el problema dual, hay que maximizar este valor de  $L^*$ , tal que el pendiente es  $\bar{\mu}$  y tiene contacto con con  $G$  en  $(\bar{y}, \bar{z})$ . Entonces la solución del dual es  $\bar{\mu}$ , y su valor es  $\bar{z}$ .



La solución del primal también es  $\bar{z}$ . En este caso, tenemos dualidad fuerte.

#### Dualidad débil



La no convexidad de  $G$  provoca el gap.

## 4.2 Problemas

1. Resuelve el primal y dual del siguiente problema

$$\begin{array}{ll} \min & (x_1 - 1)^2 + (x_2 - 1)^2 \\ \text{s.a} & x_1 + 2x_2 - 1 \leq 0 \\ & 2x_1 + x_2 - 1 \leq 0 \end{array}$$

2. Resuelve el primal y dual del siguiente problema

$$\begin{array}{ll} \min & 2x_1^2 + x_2^2 \\ \text{s.a} & x_1 + x_2 \leq 4 \\ & -x_1 + x_2 \geq -2 \end{array}$$





# Práctica 1

part.1chapter.1section.1.1section.1.2section.1.3subsection.1.3.1

<b>5</b>	<b>Práctica 1</b> .....	<b>81</b>
5.1	Introducción	
5.2	Código	
5.3	Código	
5.4	Preguntas	





## 5. Práctica 1

### 5.1 Introducción

Punto inicial  
Bounds/cotas  
Restricciones  
Función Objetivo  
Gráfica  
Multistart

### 5.2 Código

### 5.3 Código

Paquetes por importar

---

```
import numpy as np
from scipy.optimize import minimize
from scipy.optimize import Bounds
import matplotlib.pyplot as plt
```

---

Gráficas

---

```
fig = plt.figure()
plt.contourf(X, Y, Z, cmap='viridis')
plt.xlabel("x")
plt.ylabel("y")
plt.title('plot')
plt.show()
```

---

Restricciones

---

```
cons = ({'type': 'ineq', 'fun' : lambda x: -x[0]*x[1] + x[0] + x[1] -
```

```
{'type': 'ineq', 'fun' : lambda x: x[0]*x[1] + 10 }
)
```

Función Objetivo

```
def obj(x):
    return np.exp(x[0]) * (4*x[0]**2 + 2*x[1]**2 + 4*x[0]*x[1] + 2*x[1])
```

Resolver

```
res = minimize(obj, x0, method='SLSQP',
               options={'ftol': 1e-9, 'disp': False},
               bounds=bounds, constraints=cons)
```

## 5.4 Preguntas

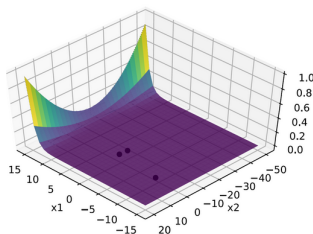
1.

$$\begin{aligned} \min \quad & e^{x_1}(4x_1^2 + 2x_2^2 + 4x_1x_2 + 2x_2 + 1) \\ \text{s.a.} \quad & x_1x_2 - x_1 - x_2 \leq -1.5 \\ & -x_1x_2 \leq 10 \end{aligned}$$

Resultados: 3 puntos óptimos.

$x_0$	$f(x^*)$	$x^*$
(0,0)	0.0235	(-9.54, 1.04)
(10, 20)	3.0607	(1.18, -1.73)
(-30, -30)	141.0364	(-1.06, -6.45)

Hay que lanzar desde distintos puntos iniciales. El cuarto solo aparece si incluimos el Jacobiano?



2.

```
# Objective function
fun = lambda x: np.exp(x[0]) * (4*x[0]**2 + ...
    ... 2*x[1]**2 + 4*x[0]*x[1] + 2*x[1] + 1)
# constraints
cons = ({'type': 'ineq', 'fun': lambda x: ...
    ... -x[0]*x[1] + x[0] + x[1] - 1.5},
        {'type': 'ineq', 'fun': lambda x: ...
    ... x[0]*x[1] + 10}
)
# bounds
bnds = ((None, None), (None, None))
bnds = ((None, None), ) * 2
# initial guesses
x0 = [(0, 0)]
# Resolver el problema
res = minimize(fun, x0, method='SLSQP', ...
    ... bounds=bnds, constraints=cons)
print("Valor optimo", res.fun)
print("Valor optimo", res.x)
```

$$\begin{aligned} \min \quad & x_1x_4(x_1 + x_2 + x_3) + x_3 \\ \text{s.a.} \quad & x_1x_2x_3x_4 \geq 25 \\ & x_1^2 + x_2^2 + x_3^2 + x_4^2 = 40 \\ & 1 \leq x_1, x_2, x_3, x_4 \leq 5 \end{aligned}$$

3.

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.a.} \quad & -x_1 \leq -0.5 \\ & -x_1 - x_2 + 1 \leq 0 \\ & -x_1^2 - x_2^2 + 1 \leq 0 \\ & -9x_1^2 - x_2^2 + 9 \leq 0 \\ & -x_1^2 - x_2 \leq 0 \\ & x_1 - x_2^2 \leq 0 \end{aligned}$$

4.

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.a.} \quad & x_1^2 + x_1x_2 + x_2^2 \leq 3 \\ & 3x_1 + 2x_2 \geq 3 \end{aligned}$$

5.

$$\begin{aligned} \min \quad & x^2 + 1 \\ \text{s.a.} \quad & (x-2)(x-4) \leq 0 \end{aligned}$$

6.

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.a.} \quad & (x_1 - 1)^2 + (x_2 - 1)^2 \leq 1 \\ & (x_1 - 1)^2 + (x_2 + 1)^2 \leq 1 \end{aligned}$$