Capítulo 1

Fundamentos de Optimización

1.1 Conceptos básicos

La teoría de optimización clásica o programación matemática está constituida por un conjunto de resultados y métodos analíticos y numéricos enfocados a encontrar e identificar al mejor candidato de entre una colección de alternativas, sin tener que enumerar y evaluar explícitamente todas esas alternativas. Un problema de optimización es, en general, un problema de decisión.

Con el fin de ilustrar de forma adecuada la estructura y composición de un problema de optimización, introduciremos a continuación un sencillo ejemplo.

Ejemplo 1.1 ***(Construcción de una caja con volumen máximo)*** *Supongamos que queremos determinar las dimensiones de una caja rectangular de forma que contenga el mayor volumen posible, pero utilizando para ello una cantidad fija de material. El problema en forma abstracta se podría plantear en los siguientes términos*

$$\begin{array}{cc} \textit{Maximizar} & \textit{Volumen de la caja} \\ \textit{sujeto a} & \textit{Área lateral fija} \end{array}$$

Con el fin de resolver este problema habrá que modelizarlo matemáticamente, es decir tendremos que expresarlo en términos matemáticos. El primer paso para modelizar un problema de optimización es identificar y definir las variables que están implicadas en dicho problema, en este caso y puesto que estamos tratando de determinar el tamaño de una caja rectangular, la opción más clara es considerar como variables sus tres dimensiones rectangulares usuales (ancho, largo, alto) y que representamos con x, y, z.

Con estas variables, la función para la que tenemos que encontrar el mejor valor será el volumen de la caja que puede expresarse como

$$V(x, y, z) = xyz$$

A continuación debemos tener en cuenta las limitaciones existentes sobre el material. Como este material se utiliza para construir las paredes de la caja, necesitaremos considerar el área lateral de la misma, y si la caja tiene tapa, dicha área será

$$A(x, y, z) = 2(xy + yz + zx)$$

Por último, teniendo en cuenta que las dimensiones de la caja no pueden ser negativas el problema puede expresarse matemáticamente como

$$\begin{array}{cc} \textit{Maximizar} & xyz \\ \textit{sujeto a} & 2(xy + yz + zx) = A \\ & x, y, z \geq 0 \end{array}$$

En este ejemplo se distinguen tres elementos fundamentales: las variables del problema, una función de esas variables y un conjunto de relaciones que deben cumplir las variables del problema. Estos elementos se repetirán en todos los problemas de optimización y se definen formalmente a continuación:

1.- *Variables de decisión:* El primer elemento clave en la formulación de problemas de optimización es la selección de las variables independientes que sean adecuadas para caracterizar los posibles diseños candidatos y las condiciones de funcionamiento del sistema. Como variables independientes se sueleng elegir aquellas que tienen un impacto significativo sobre la función objetivo.

Representaremos las variables independientes se representarán mediante vectores columna de $\mathbb{R}^n$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

o vectores fila

$$\mathbf{x}^T = (x_1, \ldots, x_n)$$

Aunque para los casos $n = 1$, 2 y 3 se emplearán las notaciones usuales de x, (x, y) y (x, y, z) respectivamente.

2.- *Restricciones:* Una vez determinadas las variables independientes, el siguiente paso es establecer, mediante ecuaciones o inecuaciones las relaciones existentes entre las variables de decisión. Estas relaciones son debidas, entre otras razones, a limitaciones en el sistema, a leyes naturales o a limitaciones tecnológicas y son las llamadas restricciones del sistema. Podemos distinguir dos tipos de restricciones:

(a) *Restricciones de igualdad:* Son ecuaciones entre las variables de la forma

$$h(\mathbf{x}) = h(x_1, \ldots, x_n) = 0$$

donde $h : A \subseteq \mathbb{R}^n \to \mathbb{R}$ es una función real de variables reales definida sobre un conjunto A de números reales.

(b) *Restricciones de desigualdad:* Son inecuaciones entre las variables de la forma

$$g(\mathbf{x}) = g(x_1, \ldots, x_n) \leq 0$$

donde $g : A \subseteq \mathbb{R}^n \to \mathbb{R}$ es una función real de variables reales definida sobre un conjunto A de números reales.

Observación 1.1 *Solamente se han considerado restricciones de dos tipos: restricciones de igualdad del forma $h(x_1, \ldots, x_n) = 0$ y restricciones de desigualdad de la forma $g(x_1, \ldots, x_n) \leq 0$, debido a que siempre es posible, mediante una simple transformación, expresar el problema en términos de esta clase de restricciones, tal y como se puede apreciar en la siguiente tabla:*

Tipo de Restricción	*Transformación*	*Nueva Restricción*
$h(x_1, \ldots, x_n) = b$	$\widehat{h} = h - b$	$\widehat{h}(x_1, \ldots, x_n) = 0$
$g(x_1, \ldots, x_n) \leq c$	$\widehat{g} = g - c$	$\widehat{g}(x_1, \ldots, x_n) \leq 0$
$g(x_1, \ldots, x_n) \geq c$	$\widehat{g} = c - g$	$\widehat{g}(x_1, \ldots, x_n) \leq 0$

Observación 1.2 *Las condiciones sobre las variables del tipo $x_i \geq 0$, $x_y \leq 0$ o similares no se incluyen dentro de este conjunto de restricciones y tienen un tratamiento particular, son restricciones en las variables o llamadas de tipo cota.*

3.- *Función objetivo:* Finalmente, el último ingrediente de un problema de optimización es la función objetivo, también llamado índice de rendimiento o criterio de elección. Este es el elemento utilizado para decidir los valores adecuados de las variables de decisión que resuelven el problema de optimización. La función objetivo permite determinar los mejores valores para las variables de decisión.

Independientemente del criterio seleccionado, dentro del contexto de la optimización matemática el adjetivo "mejor" siempre indica los valores de las variables de decisión que producen el mínimo o máximo valor (según el criterio utilizado) de la función objetivo elegida.

Algunos de estos criterios pueden ser por ejemplo de tipo económico (coste total, beneficio), de tipo tecnológico (energía mínima, máxima capacidad de carga, máxima tasa de producción) o de tipo temporal (tiempo de producción mínimo) entre otros.

1. Es importante hacer notar que en esta guía se utilizará un único criterio de optimización, no estamos interesados en encontrar una solución que, por ejemplo, minimice el coste, maximice la producción y al mismo tiempo minimice la energía utilizada. Esta simplificación es muy importante, puesto que aunque en muchas situaciones prácticas sería deseable alcanzar una solución que sea la mejor con respecto a un número de criterios diferentes: la solución ideal sería maximizar beneficios al mínimo coste. No obstante una forma de tratar objetivos que chocan entre sí, es seleccionar un criterio como primario y el resto de posibles criterios como secundarios. El criterio primario se utiliza como la función objetivo del problema de optimización, mientras que los criterios secundarios serían valores mínimos y máximos aceptables que serían tratados como restricciones del problema. Los problemas que utilizan varios criterios de búsqueda entran dentro de la llamada *programación multiobjetivo.*

Con la introducción de estos tres elementos, el objetivo de los problemas de optimización matemática está claro: Un problema de optimización consiste en la búsqueda de valores para unas determinadas variables (*variables de decisión*) de forma que, cumpliendo un conjunto de requisitos representados mediante ecuaciones y/o inecuaciones algebráicas (*restricciones*) que limitarán la elección de los valores de las variables de decisión, proporcionan el mejor valor posible para una función (*función objetivo*) que es utilizada para medir el rendimiento del sistema que se estudia. Como se ha comentado previamente, con el mejor valor se quiere indicar el mayor o el menor valor posible para la función objetivo. En resumen, buscamos valores que cumplan unas condiciones y minimicen o maximicen una función que caracteriza el sistema.

El planteamiento abstracto general para resolver problemas de este tipo es el siguiente

$$\begin{array}{cl} \text{Optimizar} & f\left(x_1, \ldots, x_n\right) \\ \text{Sujeto a} & \textit{Restricciones} \end{array}$$

donde el empleo del término **Optimizar** incluirá a ambos objetivos tanto de *Minimización* como de *Maximización.* No obstante y en relación a este aspecto, la mayoría de los planteamientos pueden hacerse con uno sólo de los objetivos, por ejemplo el de minimización, ya que un problema con objetivo de maximización se puede transformar en otro equivalente con objetivo de minimización multiplicando para ello la función objetivo por (-1) tal y como podemos comprobar en la figura 1.1. El mínimo de la función $f\left(x\right) = x^2 + 1$ se alcanza en el punto $x^* = 0$. En este punto también se alcanza el máximo

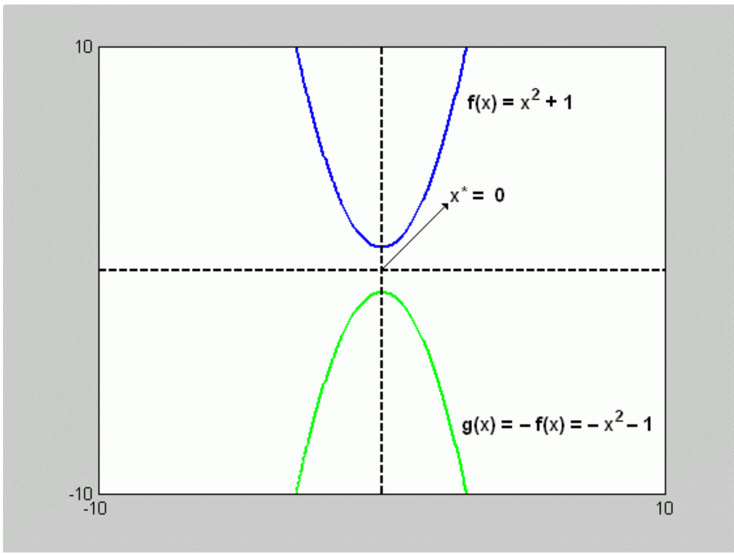

Figura 1.1: Equivalencia entre $\min f(x)$ **y** $\max g(x) = -f(x)$.

de la función opuesta $g(x) = -f(x) = -x^2 - 1$, notar que aunque el punto buscado en ambos casos es el mismo, los valores que cada función toma en dicho punto son justamente uno el opuesto del otro:

$$\begin{aligned} f(x^*) &= f(0) = 1 \\ g(x^*) &= g(0) = -1 \end{aligned}$$

Veamos algunos ejemplos de problemas de optimización.

Ejemplo 1.2 ***Distancia más corta entre dos curvas.*** *Supongamos que se quiere calcular la mínima distancia entre dos curvas de ecuaciones* $C_1 \equiv y = f(x)$ *y* $C_2 \equiv y = g(x)$ *que no se corten entre sí. El problema se resuelve considerando un punto en cada curva y utilizando la fórmula de la distancia entre dos puntos para plantear el problema como*

$$\begin{array}{ll} \textit{Minimizar} & d(P,Q) \\ \textit{sujeto a} & P \in C_1 \\ & Q \in C_2 \end{array}$$

o de forma explícita

$$\begin{array}{lc} \textit{Minimizar} & \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \\ \textit{sujeto a} & y_1 = f(x_1) \\ & y_2 = g(x_2) \end{array}$$

siendo

$$\begin{aligned} P &= (x_1, y_1) \in C_1 \\ Q &= (x_2, y_2) \in C_2 \end{aligned}$$

las coordenadas de los dos puntos. Este problema se puede extender de forma trivial a curvas en $\mathbb{R}^n$.

Ejemplo 1.3 ***Problema lineal.*** *Supongamos que queremos obtener el número de artículos que debemos fabricar de diferentes productos con coste fijo, teniendo para ello un presupuesto limitado y obteniendo a la misma vez el máximo beneficio posible. El problema podría plantearse como:*

$$\begin{array}{ll} \textit{Minimizar} & b_1x_1 + b_2x_2 + \cdots + b_nx_n \\ \\ \textit{Sujeto a} & c_1x_1 + c_2x_2 + \cdots + c_n \leq P \\ \\ & x_k \geq \quad k = 1, \ldots, n \end{array}$$

donde P es el presupuesto total disponible y los parámetros b_k y c_k para $k = 1, 2, \ldots, n$ son el beneficio y el coste, respectivamente, para cada uno de los productos y siendo x_k la cantidad de producto k que se debe fabricar.

1.2 Definiciones

En esta sección se dan las definiciones elementales relacionadas con la teoría de la optimización matemática con el objetivo de que el lector se familiarice con el lenguaje matemático utilizado.

Definición 1.1 (PPNL) *Se define el problema fundamental de la optimización estática o problema de programación no lineal (*PPNL*) al expresado como*

$$(\text{PPNL}) \quad \left\{ \begin{array}{lll} \textit{Optimizar} & f(x_1, \ldots, x_n) & \\ \\ \textit{Sujeto a} & h_i(x_1, \ldots, x_n) = 0 & i = 1, \ldots, m \\ \\ & g_j(x_1, \ldots, x_n) \leq 0 & j = 1, \ldots, p \\ \\ & (x_1, \ldots, x_n) \in A \subseteq \mathbb{R} & \end{array} \right. \tag{1.1}$$

donde $f, h_i, g_j : A \to \mathbb{R}$, o en notación vectorial como

$$(\text{PPNL}) \quad \left\{ \begin{array}{ll} \textit{Optimizar} & f(\mathbf{x}) \\ \\ \textit{Sujeto a} & \mathbf{h}(\mathbf{x}) = 0 \\ \\ & \mathbf{g}(\mathbf{x}) \leq 0 \\ \\ & \mathbf{x} = (x_1, \ldots, x_n) \in A \subseteq \mathbb{R} \end{array} \right.$$

donde ahora $\mathbf{h} : A \to \mathbb{R}^m$, y $\mathbf{g} : A \to \mathbb{R}^p$ son funciones vectoriales.

Las funciones implicadas en la definición del problema fundamental no tienen porque tener ninguna propiedad en particular, pero en nuestro caso vamos a introducir hipótesis adicionales que ayudarán a simplificar el problema; por ejemplo, supondremos de forma general que las funciones f, h_i y g_j son continuas y en la mayoría de los casos tendrán derivadas primeras y segundas también continuas.

La resolución del problema de optimización *PPNL* consistirá en primer lugar, en buscar valores para las variables de decisión x_i que cumplan las ecuaciones e inecuaciones que forman el sistema de las restricciones y en segundo lugar encontrar de entre estos valores aquel o aquellos que proporcionen el mayor (si el objetivo es maximizar) o menor (si el objetivo es minimizar) valor para la función real $f(x_1, \ldots, x_n)$.

Definición 1.2 *Se distinguen algunos casos particulares del problema general de optimización 1.1.*

1. Problemas sin restricciones: *En este tipo de problemas no hay restricciones de ningún tipo, es decir $m = p = 0$. La expresión general para estos problemas es*

$$\text{(PSR)} \begin{cases} \textit{Optimizar} & f\left(x_1, \ldots, x_n\right) \\ \\ \textit{Sujeto a} & \left(x_1, \ldots, x_n\right) \in A \end{cases} \tag{1.2}$$

Las únicas limitaciones vienen dadas por el conjunto A de $\mathbb{R}^n$ donde esté definida la función $f\left(x_1, \ldots, x_n\right)$.

2. Problemas de Lagrange o problemas sólo con restricciones de igualdad: *Son problemas de optimización con restricciones donde solamente existen restricciones de igualdad, por tanto $m \neq 0$ y $p = 0$. Son problemas de la forma*

$$\text{(PRI)} \begin{cases} \textit{Optimizar} & f\left(x_1, \ldots, x_n\right) \\ \\ \textit{Sujeto a} & h_i\left(x_1, \ldots, x_n\right) = 0 \quad i = 1, \ldots, m \\ \\ & \left(x_1, \ldots, x_n\right) \in A \end{cases} \tag{1.3}$$

No hay restricciones dadas por inecuaciones, sólo por ecuaciones.

3. Problemas unidimensionales o univariantes: *Este es un caso particular de los problemas sin restricciones en los que solamente hay una variable, es decir para $n = 1$, $m = 0$ y $p = 0$. El problema se expresa como*

$$\text{(P1D)} \begin{cases} \textit{Optimizar} & f(x) \\ \\ \textit{Sujeto a} & x \in I \subseteq \mathbb{R} \end{cases} \tag{1.4}$$

donde I es, en la mayoría de las ocasiones, un intervalo.

Definición 1.3 (Solución factible) *Diremos que $\overline{\mathbf{x}} = \left(\overline{x}_1, \ldots, \overline{x}_n\right) \in A \subseteq \mathbb{R}$ es una* solución factible *del problema* PPNL *(ecuación 1.1) si cumple todas sus restricciones, es decir*

$$\overline{\mathbf{x}} \textit{ solución factible } \Leftrightarrow \begin{cases} h_i\left(\overline{x}_1, \ldots, \overline{x}_n\right) = 0 & i = 1, \ldots, m \\ \\ g_j\left(\overline{x}_1, \ldots, \overline{x}_n\right) \leq 0 & j = 1, \ldots, p \end{cases}$$

Definición 1.4 (Conjunto factible) *Se define* región *o* conjunto factible Ω *del problema* PPNL *al conjunto de todas sus soluciones factibles*

$$\Omega = \{\overline{\mathbf{x}} \in A \subseteq \mathbb{R} : \ \overline{\mathbf{x}} \textit{ es una solución factible}\}$$

Observación 1.3 *Con estas definiciones se puede decir que el objetivo al intentar resolver el problema de optimización* PPNL *es encontrar la "mejor" de todas las soluciones factibles.*

Definición 1.5 (Mínimo global) *Diremos que $\mathbf{x}^* = \left(x_1^*, \ldots, x_n^*\right) \in \Omega \subseteq \mathbb{R}^n$ es un* mínimo global *del problema* PPNL *o que $f(\mathbf{x})$ tiene un mínimo global sobre Ω , el conjunto factible de* PPNL, *si*

$$\forall\, \overline{\mathbf{x}} \in \Omega; \quad \overline{\mathbf{x}} \neq \mathbf{x}^* \Rightarrow \quad f(\mathbf{x}^*) \leq f(\overline{\mathbf{x}})$$

El punto $\mathbf{x}^*$ será mínimo global estricto *si la desigualdad es estricta*

$$\forall\, \overline{\mathbf{x}} \in \Omega; \quad \overline{\mathbf{x}} \neq \mathbf{x}^* \Rightarrow \quad f(\mathbf{x}^*) < f(\overline{\mathbf{x}}).$$

Esta definición implica que no hay en Ω un punto en el que la función tome un valor menor que el que toma en $\mathbf{x}^*$.

Definición 1.6 (Máximo global) *Diremos que* $\mathbf{x}^* = (x_1^*, \ldots, x_n^*) \in \Omega \subseteq \mathbb{R}^n$ *es un* máximo global *del problema* PPNL *o que* $f(x)$ *tiene un máximo global sobre* Ω*, el conjunto factible de* PPNL, *si*

$$\forall\, \overline{\mathbf{x}} \in \Omega; \quad \overline{\mathbf{x}} \neq \mathbf{x}^* \Rightarrow \quad f(\mathbf{x}^*) \geq f(\overline{\mathbf{x}})$$

El punto $\mathbf{x}^*$ *será* máximo global estricto *si la desigualdad es estricta*

$$\forall\, \overline{\mathbf{x}} \in \Omega; \quad \overline{\mathbf{x}} \neq \mathbf{x}^* \Rightarrow \quad f(\mathbf{x}^*) > f(\overline{\mathbf{x}})$$

Esta definición implica que no hay en Ω un punto en el que la función tome un valor mayor que el que toma en $\mathbf{x}^*$.

Observación 1.4 *Los máximos y mínimos globales de un problema de optimización se denominan* extremos globales.

Definición 1.7 (Solución óptima) *Diremos que* $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ *es una* solución óptima *del problema* PPNL *o que* $f(\mathbf{x})$ *tiene un óptimo en* $\mathbf{x}^*$ *sobre el conjunto factible* Ω *si ocurre alguna de estas dos situaciones*

1. $\mathbf{x}^*$ *es un mínimo global del problema* PPNL *y el objetivo del problema es minimizar.*
2. $\mathbf{x}^*$ *es un máximo global del problema* PPNL *y el objetivo del problema es maximizar.*

Definición 1.8 (Valor óptimo) *Si* $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ *es una solución óptima del problema* PPNL, *entonces se define el* valor óptimo *como el valor de la función objetivo en la solución óptima, es decir, si* $\mathbf{x}^*$ *es una solución óptima del problema* PPNL, *entonces* $f\left(\mathbf{x}^{`*}\right)$ *es el valor óptimo.*

Resolver un problema de optimización es encontrar, si existen, sus soluciones óptimas, es decir los extremos globales de la función objetivo sobre el conjunto factible. Desde el punto de vista práctico y computacional en algunas ocasiones bastará con obtener los llamados extremos locales que se definen a continuación.

Definición 1.9 (Mínimo local) *Consideremos el problema de optimización* PPNL *y sea* Ω *su conjunto factible. Diremos que* $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ *es un* mínimo local o relativo *de* $f(\mathbf{x})$ *en* Ω *si y sólo si*

$$\exists\, \epsilon > 0 \textit{ tal que } \forall\, \overline{\mathbf{x}} \in \Omega; \quad \overline{\mathbf{x}} \neq \mathbf{x}^*; \quad \|\overline{\mathbf{x}} - \mathbf{x}^*\| < \epsilon \Longrightarrow f(\mathbf{x}^*) \leq f(\overline{\mathbf{x}})$$

El punto $\mathbf{x}^*$ *será un* mínimo local estricto *de* $f(x)$ *en* Ω *si la desigualdad es estricta*

$$\exists\, \epsilon > 0 \textit{ tal que } \forall\, \overline{\mathbf{x}} \in \Omega; \quad \overline{\mathbf{x}} \neq \mathbf{x}^*; \quad \|\overline{\mathbf{x}} - \mathbf{x}^*\| < \epsilon \Longrightarrow f(\mathbf{x}^*) < f(\overline{\mathbf{x}}).$$

Definición 1.10 (Máximo local) *Consideremos el problema general de optimización* PPNL *y sea* Ω *su conjunto factible. Diremos que* $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ *es un* máximo local o relativo *de* $f(\mathbf{x})$ *en* Ω *si y sólo si*

$$\exists\, \epsilon > 0 \textit{ tal que } \forall\, \overline{\mathbf{x}} \in \Omega; \quad \overline{\mathbf{x}} \neq \mathbf{x}^*; \quad \|\overline{\mathbf{x}} - \mathbf{x}^*\| < \epsilon \Longrightarrow f(\mathbf{x}^*) \geq f(\overline{\mathbf{x}})$$

El punto $\mathbf{x}^*$ *será un* máximo local estricto *de* $f(x)$ *en* Ω *si la desigualdad es estricta*

$$\exists\, \epsilon > 0 \textit{ tal que } \forall\, \overline{\mathbf{x}} \in \Omega; \quad \overline{\mathbf{x}} \neq \mathbf{x}^*; \quad \|\overline{\mathbf{x}} - \mathbf{x}^*\| < \epsilon \Longrightarrow f(\mathbf{x}^*) > f(\overline{\mathbf{x}}).$$

Observación 1.5 *Los máximos y mínimos locales de los problemas de optimización también se denominan* extremos locales o relativos.

A diferencia de los extremos globales que afectan a todo el conjunto factible Ω, los extremos locales afectan a cierto entorno a su alrededor.

La teoría inicial asociada a la optimización está orientada a la obtención de condiciones necesarias y suficientes para que un punto sea óptimo. Como veremos en los temas correspondientes, esta teoría incluye el teorema de los multiplicadores de Lagrange y el teorema de Karush-Kuhn-Tucker. Por otra parte, también es interesante conocer no sólo si un punto es o no óptimo desde el punto de vista teórico, sino también cómo encontrar esos óptimos desde el punto de vista práctico. Teniendo esto en cuenta, al considerar problemas de optimización se plantean dos cuestiones:

1. **Cuestión estática:** ¿Cómo podemos determinar si un punto $\mathbf{x}^*$ es o no la solución óptima de un problema de optimización? ¿Qué condiciones deberían cumplir las funciones $f(\mathbf{x})$, $h_i(\mathbf{x})$ y $g_j(\mathbf{x})$ para que un problema *PPNL* tenga solución? ¿Qué condiciones debe cumplir el punto $\mathbf{x}^*$?

2. **Cuestión dinámica:** Si $\overline{\mathbf{x}}$ no es el punto óptimo, entonces ¿cómo podemos encontrar una solución óptima $\mathbf{x}^*$, utilizando la información de la función en $\overline{\mathbf{x}}$?

Mientras que con la primera cuestión se trata de determinar condiciones necesarias y/o suficientes para que un punto sea o no una solución óptima, en la segunda de las cuestiones se consideran los métodos numéricos adecuados para conseguir encontrar esas soluciones óptimas.

El resultado principal utilizado para conocer si un problema de optimización tiene solución es el *teorema de Weierstrass*, que recordamos a continuación dentro del contexto de la optimización matemática.

Teorema 1.1 (Teorema de Weierstrass) *Sea $f(\mathbf{x})$ una función continua definida sobre un conjunto compacto (cerrado y acotado) $K \subseteq \mathbb{R}^n$. Entonces el problema de optimización*

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

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

$$\exists \quad \mathbf{x}^*_{\min}, \mathbf{x}^*_{\max} \in K : \begin{cases} f(\mathbf{x}^*_{\min}) = \min\limits_{\mathbf{x} \in K} f(\mathbf{x}) \\ f(\mathbf{x}^*_{\max}) = \max\limits_{\mathbf{x} \in K} f(\mathbf{x}) \end{cases}$$

Este es un resultado importante a tener en cuenta en la resolución de problemas de optimización, sin embargo el teorema no nos proporciona un método para la localización de las soluciones, solamente de su existencia en determinadas condiciones. Desde el punto de vista de las aplicaciones, lo interesante es caracterizar los puntos solución y diseñar un método efectivo para su cálculo.

Finalmente, apuntar que un problema de optimización puede tener solución única como en el siguiente planteamiento

$$\begin{cases} \text{Optimizar} & x^2 \\ \text{Sujeto a} & x \in [0,1] \end{cases},$$

no tener ninguna solución como en el problema

$$\begin{cases} \text{Optimizar} & \dfrac{1}{x} \\ \\ \text{Sujeto a} & x \in (0,1) \end{cases},$$

o tener más de una solución, como en el problema

$$\begin{cases} \text{Optimizar} & \operatorname{sen}(x) \\ \\ \text{Sujeto a} & x \in \mathbb{R} \end{cases}$$

en el que hay incluso infinitas soluciones óptimas, tanto de mínimo ($\frac{3\pi}{2}+2k\pi,\ k \in \mathbb{Z}$) como de máximo ($\frac{\pi}{2}+2k\pi,\ k \in \mathbb{Z}$).

1.3 Conjuntos Convexos

El concepto de convexidad es de gran importancia en el estudio de los problemas de optimización desde el punto de vista de la aplicación práctica, puesto que en algunos casos, bajo condiciones de convexidad, se puede garantizar que un extremo local de un problema es realmente un extremo global y por tanto la solución óptima del problema buscada.

Se describen en esta sección algunos conceptos básicos de convexidad útiles para el desarrollo de la programación matemática y aunque es posible definirlos en el ámbito de cualquier espacio topológico, en lo sucesivo consideraremos el espacio vectorial $\mathbb{R}^n$.

Definición 1.11 (Segmento lineal) *Dados dos puntos* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *el* segmento lineal cerrado *que une* $\mathbf{x}$ *con* $\mathbf{y}$ *es el conjunto definido por*

$$[\mathbf{x}, \mathbf{y}] = \{\lambda \mathbf{x} + (1-\lambda)\,\mathbf{y} \in \mathbb{R}^n : 0 \leq \lambda \leq 1\}$$

Del mismo modo, el segmento lineal abierto *que une* $\mathbf{x}$ *con* $\mathbf{y}$ *es el conjunto*

$$(\mathbf{x}, \mathbf{y}) = \{\lambda \mathbf{x} + (1-\lambda)\,\mathbf{y} \in \mathbb{R}^n : 0 < \lambda < 1\}$$

Definición 1.12 (Conjunto convexo) *Sea* $\Omega \subset \mathbb{R}^n$ *entonces*

$$\Omega \textit{ es convexo } \Longleftrightarrow \forall\ \mathbf{x}, \mathbf{y} \in \Omega \Rightarrow [\mathbf{x}, \mathbf{y}] \subseteq \Omega$$

Esta definición se interpreta de forma que un conjunto será convexo si el segmento lineal cerrado que une cualquier par de puntos del conjunto está contenido en dicho conjunto. La figura 1.2 representa algunos conjuntos convexos y otros no convexos de $\mathbb{R}^2$.

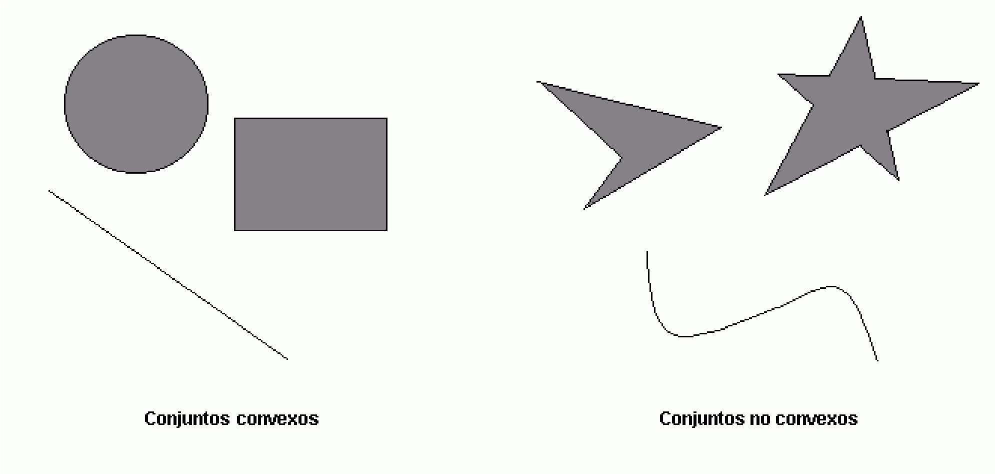

Figura 1.2: Convexidad en $\mathbb{R}^2$.

Por convenio el conjunto vacío $\emptyset$ es un conjunto convexo.

Uno de los tipos más importantes de conjunto convexo es el *hiperplano*, que definimos a continuación.

Definición 1.13 (Hiperplano) *Sea* $\mathbf{a} \in \mathbb{R}^n$ *con* $\mathbf{a} \neq 0$ *y* $b \in \mathbb{R}$. *Un* hiperplano H *en* $\mathbb{R}^n$ *es un conjunto definido como*

$$H = \left\{ \mathbf{x} \in \mathbb{R}^n : \ \mathbf{a}^T \mathbf{x} = b \right\}$$

El vector $\mathbf{a}$ *es el llamado vector* normal *al hiperplano. La expresión* $\mathbf{a}^T\mathbf{x}$ *es el producto escalar de ambos vectores*

$$\mathbf{a}^T \mathbf{x} = \ a_1 x_1 + \cdots + a_n x_n$$

Definición 1.14 (Semiespacios) *Sea* $\mathbf{a} \in \mathbb{R}^n$ *con* $\mathbf{a} \neq \mathbf{0}$ *y* $b \in \mathbb{R}$. *Sea* H *el hiperplano construido a partir de* $\mathbf{a}$ *y* b *entonces definimos los* semiespacios cerrados positivos *y* negativos *asociados a* H*, respectivamente a los conjuntos*

$$\begin{aligned} H_+ &= \left\{ \mathbf{x} \in \mathbb{R}^n : \ \mathbf{a}^T \mathbf{x} \geq b \right\} \\ H_- &= \left\{ \mathbf{x} \in \mathbb{R}^n : \ \mathbf{a}^T \mathbf{x} \leq b \right\} \end{aligned}$$

y semiespacios abiertos positivos *y* negativos *asociados a* H *a los conjuntos definidos como*

$$\begin{aligned} \overset{\circ}{H}_+ &= \left\{ \mathbf{x} \in \mathbb{R}^n : \ \mathbf{a}^T \mathbf{x} > b \right\} \\ \overset{\circ}{H}_- &= \left\{ \mathbf{x} \in \mathbb{R}^n | : \ \mathbf{a}^T \mathbf{x} < b \right\} \end{aligned}$$

El siguiente lema es una consecuencia inmediata de la definición de convexidad y establece que la intersección de dos conjuntos convexos es convexa y que la suma algebráica de dos conjuntos convexos también es convexa. Su demostración es muy sencilla utilizando la propia definición de conjunto convexo y se deja como ejercicio.

Lema 1.2 *Sean* Ω_1, $\Omega_2 \subseteq \mathbb{R}^n$ *dos conjuntos convexos entonces*

1. *$\Omega_1 \cap \Omega_2$ es convexo.*

2. *$\Omega_1 + \Omega_2 = \{\mathbf{x} + \mathbf{y} \in \mathbb{R}^n : \mathbf{x} \in \Omega_1,\ \mathbf{y} \in \Omega_2\}$ es convexo.*

3. *$\Omega_1 - \Omega_2 = \{\mathbf{x} - \mathbf{y} \in \mathbb{R}^n : \mathbf{x} \in \Omega_1,\ \mathbf{y} \in \Omega_2\}$ es convexo.*

Observación 1.6 *Es posible extender mediante inducción la propiedad* 1 *del lema 1.2 a una intersección cualquiera de conjuntos convexos, es decir si $\{\Omega_n\}_{n=1}^{\infty} \subseteq \mathbb{R}^n$ es una familia de conjuntos convexos de $\mathbb{R}^n$, entonces el conjunto intersección $\bigcap_{n=1}^{\infty} \Omega_n$ es de nuevo un conjunto convexo.*

Si tenemos en cuenta por una parte que los semiespacios de $\mathbb{R}^n$ son conjuntos convexos y por otra utilizamos los resultados del lema 1.2 se deduce fácilmente que los conjuntos definidos de la forma

$$\Omega = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \begin{array}{c} a_{11}x_1 + \cdots + a_{1n}x_n \leq b_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \leq b_m \end{array} \right\}$$

o en de forma más compacta en notación matricial

$$\Omega = \{\mathbf{x} \in \mathbb{R}^n : \quad \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$$

donde $\mathbf{A} = (a_{ij})$ es una matriz $m \times n$ en $\mathbb{R}$ y $\mathbf{b} \in \mathbb{R}^m$ es un vector; son convexos puesto que

$$\Omega = \Omega_1 \cap \cdots \cap \Omega_m$$

siendo

$$\Omega_k = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ a_{k1}x_1 + \cdots + a_{kn}x_n \leq b_k\}\,; \quad k = 1, \ldots, m$$

semiespacios cerrados negativos, que son conjuntos convexos.

Ejemplo 1.4 *El conjunto definido como $S = \{\mathbf{x} \in \mathbb{R}^n$: $\mathbf{x}$ es solución óptima de $P\}$ siendo P el problema*

$$P = \left\{ \begin{array}{ll} \textit{Minimizar} & \mathbf{c}^T\mathbf{x} \\ \\ \textit{Sujeto a} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ \\ & \mathbf{x} \geq 0 \end{array} \right.$$

es un conjunto convexo.

Solución: Para demostrar la convexidad de S utilizaremos la definición. Distinguimos tres caso:

1. Caso I: El problema P no tiene solución, en ese caso S es el conjunto vacío $\emptyset$, que por definición es un conjunto convexo.

2. Caso II: El problema P tiene solución única: $S = \{\mathbf{x}^*\}$. En este caso el único segmento lineal que se puede construir es
$$\lambda\mathbf{x}^* + (1 - \lambda)\,\mathbf{x}^* = \mathbf{x}^* \in S$$
(en este caso incluso independientemente de si λ está en el intervalo $[0, 1]$ o no).

3. Caso III: El problema P tiene más de una solución. En este caso S tiene más de un elemento y podemos suponer que existen $\mathbf{x}^1$ y $\mathbf{x}^2 \in S$ que por ser soluciones del problema P serán puntos factibles y por tanto cumplirán las restricciones del problema

$$\begin{array}{c} \mathbf{A}\mathbf{x}^1 = \mathbf{b} \\ \mathbf{A}\mathbf{x}^2 = \mathbf{b} \\ \\ \mathbf{x}^1, \mathbf{x}^2 \geq 0 \end{array}$$

Como además son mínimos de P; si llamamos f^* al valor óptimo se tiene

$$f^* = \mathbf{c}^T\mathbf{x}^1 = \mathbf{c}^T\mathbf{x}^2$$

Ahora hay que comprobar que los elementos del segmento lineal que une $\mathbf{x}^1$ con $\mathbf{x}^2$ también están en S es decir son soluciones de P. Para ello tomamos $\lambda \in [0,1]$ y consideremos el punto $\mathbf{x}^3$ definido por

$$\mathbf{x}^3 = \lambda\mathbf{x}^1 + (1-\lambda)\,\mathbf{x}^2$$

Comprobamos su factibilidad operando directamente. Por una parte $\mathbf{x}^3$ es factible ya que por una parte cumple el sistema de ecuaciones

$$\mathbf{A}\mathbf{x}^3 = \mathbf{A}\left(\lambda\mathbf{x}^1 + (1-\lambda)\,\mathbf{x}^2\right) = \lambda\mathbf{A}\mathbf{x}^1 + (1-\lambda)\,\mathbf{A}\mathbf{x}^2 = \lambda\mathbf{b} + (1-\lambda)\,\mathbf{b} = \mathbf{b},$$

y por otra parte, como $\lambda \in [0,1]$

$$0 \leq \lambda \leq 1 \Rightarrow \left\{ \begin{array}{c} \lambda \geq 0 \\ \\ 1-\lambda \geq 0 \end{array} \right.$$

de donde al ser $\mathbf{x}^1, \mathbf{x}^2 \geq 0$ se obtiene

$$\mathbf{x}^3 = \lambda\mathbf{x}^1 + (1-\lambda)\,\mathbf{x}^2 \geq 0$$

ya que todos los sumandos son positivos. De esta forma $\mathbf{x}^3$ es factible porque cumple las restricciones del problema.

Si ahora evaluamos la función objetivo en $\mathbf{x}^3$

$$f\left(\mathbf{x}^3\right) = \mathbf{c}^T\mathbf{x}^3 = \mathbf{c}^T\left(\lambda\mathbf{x}^1 + (1-\lambda)\,\mathbf{x}^2\right) = \lambda\mathbf{c}^T\mathbf{x}^1 + (1-\lambda)\,\mathbf{c}^T\mathbf{x}^2 = \lambda f^* + (1-\lambda)\,f^* = f^*$$

y por tanto se deduce que $\mathbf{x}^3$ también es solución del problema P.

Definición 1.15 (Punto extremo) *Sea $\Omega \subseteq \mathbb{R}^n$ un conjunto convexo no vacío. Se dice que el punto $\mathbf{x} \in \Omega$ es un* vértice o punto extremo *de $\Omega \Longleftrightarrow$*

$$Si\ \mathbf{x} = \lambda\mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 \ para\ algún\ \lambda \in [0,1]\ \ y\ \mathbf{x}^1, \mathbf{x}^2 \in \Omega \Longrightarrow \mathbf{x}^1 = \mathbf{x}^2 = \mathbf{x}$$

Por ejemplo el conjunto convexo

$$\Omega = \left\{ \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \middle| \quad 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 2 \right\}$$

tiene 4 puntos extremos dados por $P_1 = (0,0)$, $P_2 = (0,2)$, $P_3 = (2,0)$ y $P_4 = (2,2)$ (ver figura 1.3).

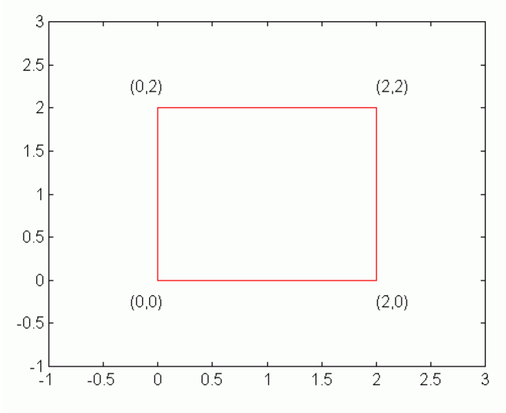

Figura 1.3: Puntos extremos.

El conjunto de los puntos extremos de un conjunto convexo puede ser vacío, por ejemplo una bola abierta de $\mathbb{R}^n$, contener una cantidad finita de elementos, como en la figura 1.3 o tener una cantidad infinita de elementos, como una bola cerrada de $\mathbb{R}^n$.

1.4 Funciones convexas

En esta sección se proporciona la definición de función convexa y se presentan alguna de sus propiedades más importantes, sobre todo aquellas que pueden utilizarse para resolver problemas de optimización.

Definición 1.16 (Función convexa) *Diremos que la función $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, con Ω un conjunto convexo no vacío, es* convexa *sobre $\Omega \iff$*

$$\forall \mathbf{x}, \mathbf{y} \in \Omega, \lambda \in [0,1] \Rightarrow f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

Se dice que f estrictamente convexa $\iff$

$$\forall \mathbf{x}, \mathbf{y} \in \Omega, \lambda \in [0,1] \Rightarrow f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

Observación 1.7 *Si definimos $\mu = 1-\lambda$, entonces la definición puede ponerse como*

$$f \text{ es convexa} \Leftrightarrow f(\lambda \mathbf{x} + \mu \mathbf{y}) \leq \lambda f(\mathbf{x}) + \mu f(\mathbf{y}), \quad \forall \lambda, \mu \geq 0 \quad \text{con } \lambda + \mu = 1$$

Y del mismo modo para función estrictamente convexa.

Definición 1.17 (Función cóncava) *Diremos que la función $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, con Ω conjunto convexo no vacío es* cóncava *sobre $\Omega \iff g = -f$ es convexa.*

Esta definición equivale a decir que

$$f \text{ es cóncava} \Leftrightarrow \forall \mathbf{x}, \mathbf{y} \in \Omega, \lambda \in [0,1] \Rightarrow f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

Definición 1.18 *Se dice que f* estrictamente cóncava $\iff g = -f$ *es estrictamente convexa.*

Proposición 1.3 *Si $f_1(\mathbf{x})$ y $f_2(\mathbf{x})$ son dos funciones convexas definidas sobre un conjunto convexo no vacío $\Omega \subseteq \mathbb{R}^n \Longrightarrow f(\mathbf{x}) = (f_1 + f_2)(\mathbf{x})$ es convexa sobre Ω.*

Demostración: Por definición de $f = f_1 + f_2$ se tiene

$$f(\mathbf{x}) = (f_1 + f_2)(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$$

Sean $\mathbf{x}, \mathbf{y} \in \Omega$ y $\lambda \in [0, 1]$

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = f_1(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + f_2(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})$$

Puesto que tanto f_1 como f_2 son convexas sobre Ω se cumplen de forma simultanea las siguientes desigualdades

$$\begin{aligned} f_1(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &\leq \lambda f_1(\mathbf{x}) + (1-\lambda) f_1(\mathbf{y}) \\ f_2(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &\leq \lambda f_2(\mathbf{x}) + (1-\lambda) f_2(\mathbf{y}) \end{aligned}$$

y sumando ambas expresiones

$$f_1(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + f_2(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda(f_1(\mathbf{x}) + f_2(\mathbf{x})) + (1-\lambda)(f_1(\mathbf{y}) + f_2(\mathbf{y}))$$

es decir

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

por tanto $f(\mathbf{x})$ es convexa.

Proposición 1.4 *Si $f(\mathbf{x})$ es convexa sobre $\Omega \subseteq \mathbb{R}^n$, siendo Ω un conjunto convexo no vacío, entonces $\forall \alpha \geq 0$, la función $(\alpha f)(\mathbf{x})$ definida por*

$$(\alpha f)(\mathbf{x}) = \alpha f(\mathbf{x})$$

es convexa sobre Ω.

Demostración: Como $f(\mathbf{x})$ es convexa sobre Ω se cumple

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

si ahora multiplicamos ambos lados de la desigualdad por $\alpha > 0$, el sentido de esta no cambia y tendremos

$$\alpha(f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})) \leq \alpha(\lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})) = \lambda\alpha f(\mathbf{x}) + (1-\lambda)\alpha f(\mathbf{x})$$

lo que demuestra la convexidad de αf.

Observación 1.8 *Si $\alpha < 0$, entonces la función αf sería cóncava sobre Ω.*

Proposición 1.5 *Sea f convexa en $\Omega \subseteq \mathbb{R}^n$, con Ω conjunto convexo no vacío y sea $c \in \mathbb{R}$, entonces el conjunto de nivel Γ_c definido por*

$$\Gamma_c = \{\mathbf{x} \in \Omega : f(\mathbf{x}) \leq c\}$$

es convexo $\forall c \in \mathbb{R}$.

Demostración: Hay que demostrar que $\forall\ \mathbf{x}, \mathbf{y} \in \Gamma_c$ y $\lambda \in [0,1]$, debe ocurrir $\lambda\mathbf{x} + (1-\lambda)\,\mathbf{y} \in \Gamma_c$. Utilizando la definición de Γ_c es equivalente a probar

$$\forall\ \mathbf{x}, \mathbf{y} \text{ con } f(\mathbf{x}) \leq c \text{ y } f(\mathbf{y}) \leq c \implies f(\lambda\mathbf{x} + (1-\lambda)\,\mathbf{y}) \leq c \quad \forall\lambda \in [0,1]$$

Como $f(\mathbf{x})$ es convexa se cumple

$$f(\lambda\mathbf{x} + (1-\lambda)\,\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)\, f(\mathbf{y})$$

Por otra parte como $\lambda \in [0,1]$ se cumple

$$0 \leq \lambda \leq 1 \Longrightarrow \begin{cases} 0 \leq \lambda \\ \\ \lambda \leq 1 \Rightarrow 0 \leq 1-\lambda \end{cases}$$

y por tanto

$$\begin{aligned} f(\mathbf{x}) &\leq c \Longrightarrow \lambda f(\mathbf{x}) \leq \lambda c \\ \\ f(\mathbf{y}) &\leq c \Longrightarrow (1-\lambda)\, f(\mathbf{x}) \leq (1-\lambda)\, c \end{aligned}$$

Al sumar ambas desigualdades se obtiene

$$\lambda f(\mathbf{x}) + (1-\lambda)\, f(\mathbf{x}) \leq \lambda c + (1-\lambda)\, c = c$$

y por tanto

$$f(\lambda\mathbf{x} + (1-\lambda)\,\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)\, f(\mathbf{x}) \leq c$$

como tratábamos de probar. ■

La caracterización de funciones convexas mediante su definición es, en general, muy difícil de aplicar en la práctica. Para comprobar si una función es o no convexa es necesario encontrar otras caracterizaciones más sencillas de aplicar. Los siguientes resultados proporcionan esas caracterizaciones para funciones convexas diferenciables en términos del gradiente y del Hessiano.

Proposición 1.6 (Caracterización de primer orden para funciones convexas) *Sea $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, con Ω un conjunto convexo y $f(\mathbf{x}) \in \mathcal{C}^1(\Omega)$, es decir una función derivable en Ω con derivadas parciales continuas, entonces*

$$f(\mathbf{x}) \text{ es convexa sobre } \Omega \Longleftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}) \quad \forall\mathbf{x}, \mathbf{y} \in \Omega$$

$$f(\mathbf{x}) \text{ es estrictamente convexa sobre } \Omega \Longleftrightarrow f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}) \quad \forall\mathbf{x}, \mathbf{y} \in \Omega$$

Demostración: Demostramos la proposición para el caso en el que $f(\mathbf{x})$ sea convexa y se procedería del mismo modo si $f(\mathbf{x})$ fuera estrictamente convexa.

"$\Longrightarrow$" Supongamos en primer lugar que $f(\mathbf{x})$ es convexa, entonces tomando cualquier valor $\lambda \in [0,1]$ y $\mathbf{x}, \mathbf{y} \in \Omega$, se tiene

$$f(\lambda\mathbf{y} + (1-\lambda)\mathbf{x}) \leq \lambda f(\mathbf{y}) + (1-\lambda) f(\mathbf{x})$$

Se puede asumir el hecho de que $\lambda \neq 0$ y podremos entonces dividir la inecuación por este valor

$$\frac{f(\mathbf{x} + \lambda(\mathbf{y}-\mathbf{x})) - f(\mathbf{x})}{\lambda} \leq f(\mathbf{y}) - f(\mathbf{x}) \quad \forall\lambda \in (0,1]$$

Si ahora $\lambda \to 0$ y tenemos en cuenta que la función $f(\mathbf{x})$ es derivable en Ω, entonces

$$\lim_{\lambda \to 0} \frac{f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\lambda} = \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

y por tanto

$$\nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$$

"$\Longleftarrow$" Supongamos ahora que es cierta la siguiente desigualdad

$$\nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}) \; \forall \mathbf{x}, \mathbf{y} \in \Omega$$

entonces por ser Ω un conjunto convexo

$$\forall\, \mathbf{x}^1, \mathbf{x}^2 \in \Omega \text{ y } \lambda \in [0,1] \Longrightarrow \mathbf{x} = \lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 \in \Omega$$

Si utilizamos la desigualdad para $\mathbf{x}$ e $\mathbf{y} = \mathbf{x}^1$ por una parte y para $\mathbf{x}$ e $\mathbf{y} = \mathbf{x}^2$ por otra obtenemos

$$\begin{aligned} \mathbf{y} = \mathbf{x}^1 \quad &\Longrightarrow \quad f(\mathbf{x}^1) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{x}^1 - \mathbf{x}) \\ \mathbf{y} = \mathbf{x}^2 \quad &\Longrightarrow \quad f(\mathbf{x}^2) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{x}^2 - \mathbf{x}) \end{aligned}$$

Si multiplicamos por λ la primera desigualdad y por $(1-\lambda)$ la segunda y sumamos ambas expresiones

$$\begin{array}{rcl} \lambda f\left(\mathbf{x}^1\right) & \geq & \lambda f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})(\mathbf{x}^1 - \mathbf{x}) \\ (1-\lambda) f\left(\mathbf{x}^2\right) & \geq & (1-\lambda) f(\mathbf{x}) + (1-\lambda)\nabla f(\mathbf{x})(\mathbf{x}^2 - \mathbf{x}) \\ \hline \lambda f(\mathbf{x}^1) + (1-\lambda) f(\mathbf{x}^2) & \geq & f(\mathbf{x}) + \nabla f(\mathbf{x})(\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 - \mathbf{x}) \\ & \geq & f(\mathbf{x}) = f(\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2) \end{array}$$

lo que demuestra la convexidad de $f(\mathbf{x})$. ■

Proposición 1.7 (Caracterización de segundo orden para funciones convexas) *Sea $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, con Ω un conjunto abierto convexo no vacío y $f(\mathbf{x}) \in \mathcal{C}^2(\Omega)$ entonces:*

$$f(\mathbf{x}) \;\; \textit{es convexa en } \Omega \Longleftrightarrow \mathbf{H}f(\mathbf{x}) \;\; \textit{es semidefinido positivo } \forall \mathbf{x} \in \Omega$$

siendo

$$\mathbf{H}f(\mathbf{x}) = \left[\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \right)_{i,j=1}^{n} \right]$$

la matriz hessiana asociada a $f(\mathbf{x})$.

Demostración: Demostramos la doble implicación.

"$\Longrightarrow$" Supongamos que $f(\mathbf{x})$ es convexa en Ω y sea $\mathbf{x} \in \Omega$. Demostrar que $\mathbf{H}f(\mathbf{x})$ es semidefinida positiva en Ω equivale a demostrar que $\forall\, \mathbf{d} \in \mathbb{R}^n$ ocurre $\mathbf{d}^T \mathbf{H} f(\mathbf{x}) \mathbf{d} \geq 0$.

Puesto que $\Omega \neq \emptyset$ y abierto, entonces para $\mathbf{d} \in \mathbb{R}^n$, ocurre $\mathbf{y} = \mathbf{x} + \lambda \mathbf{d} \in \Omega$ tomando $|\lambda| \neq 0$ y suficientemente pequeño. Si utilizamos la proposición 1.6 de caracterización de primer orden de funciones convexas por una parte y el teorema de Taylor por otra, se obtienen las siguientes expresiones:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})\,(\mathbf{y}-\mathbf{x}) \Leftrightarrow f(\mathbf{x}+\lambda\mathbf{d}) \geq f(\mathbf{x}) + \lambda\nabla f(\mathbf{x})\mathbf{d}$$

$$f(\mathbf{x}+\lambda\mathbf{d}) = f(\mathbf{x}) + \lambda\nabla f(\mathbf{x})^T\mathbf{d} + \tfrac{1}{2}\lambda^2\mathbf{d}^T\mathbf{H}f(\mathbf{x})\mathbf{d} + o(\lambda)$$

donde $o(\lambda) = \lambda^2||\mathbf{d}||^2\alpha(\mathbf{x},\lambda\mathbf{d})$ es una función que converge a 0 cuando λ tiende a 0.

Si se restan ambas expresiones obtenemos

$$\frac{1}{2}\lambda^2\mathbf{d}^T\mathbf{H}f(\mathbf{x})\mathbf{d} + \lambda^2||\mathbf{d}||^2\alpha(\mathbf{x},\lambda\mathbf{d}) \geq 0$$

A continuación se divide por λ^2 y se toman límites cuando $\lambda \to 0$. El resultado que se obtiene es que $\mathbf{d}^T\mathbf{H}f(\mathbf{x})\mathbf{d} \geq 0$.

“$\Longleftarrow$” Recíprocamente si suponemos ahora que la matriz Hessiana $\mathbf{H}f\,(\mathbf{x})$ es semidefinida positiva en cada punto de Ω y consideramos $\mathbf{x}$ e $\mathbf{y}$ en Ω, entonces por el teorema del valor medio obtenemos:

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^T\mathbf{H}f(\mathbf{x}^*)(\mathbf{y}-\mathbf{x})$$

donde $\mathbf{x}^* = \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \Omega$ con $\lambda \in (0,1)$. Como $\mathbf{H}f(\mathbf{x}^*)$ semidefinida positiva, se cumple que $(\mathbf{y}-\mathbf{x})^T\mathbf{H}f(\mathbf{x}^*)(\mathbf{y}-\mathbf{x}) \geq 0$ y concluimos que

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^T\mathbf{H}f(\mathbf{x}^*)(\mathbf{y}-\mathbf{x}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x})$$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \lambda\nabla f(\mathbf{x})\mathbf{x} \quad \forall\mathbf{x}^*, \mathbf{x} \in \Omega$$

que resulta ser la caracterización de primer orden para funciones convexas de la proposición 1.6 y por tanto $f\,(x)$ es convexa en Ω $\blacksquare$.

La matriz Hessiana de f es la generalización al espacio $\mathbb{R}^n$ del concepto de curvatura de una función y de forma análoga, la definición positiva del Hessiano es la generalización de curvatura positiva. Las funciones convexas tienen curvatura positiva (o al menos no negativa) en todas las direcciones.

1.5 Optimización de funciones convexas

En esta sección consideraremos el problema de minimizar y maximizar una función convexa sobre un conjunto convexo.

Debido a la correspondencia entre funciones cóncavas y convexas todos los resultados se presentan de forma equivalente para ambos tipos de funciones, por ello en cada teorema y entre corchetes se muestra el resultado alternativo.

Teorema 1.8 *Sea $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ con Ω un conjunto convexo. Si $f\,(\mathbf{x})$ es convexa [cóncava] $\Longrightarrow$ El conjunto Γ donde $f\,(\mathbf{x})$ alcanza su mínimo [máximo] es convexo y cualquier mínimo [máximo] local de $f\,(\mathbf{x})$ es mínimo [máximo] global.*

Demostración: Definimos el conjunto Γ como

$$\Gamma = \{\mathbf{x}^* \in \Omega | \ f(\mathbf{x}^*) = \min_{x \in \Omega} f(x)\}$$

Si $\Gamma = \emptyset$, es decir si $f(\mathbf{x})$ no tiene mínimos relativos, entonces el teorema se cumple por la convexidad del conjunto vacío.

Supongamos ahora $\Gamma \neq \emptyset$ y por tanto existe $\mathbf{x}^* \in \Gamma$. Si $\mathbf{x}^*$ es el único elemento de Γ, entonces es convexo puesto que la única combinación con elementos de Γ que se puede hacer $\forall \lambda \in [0,1]$ es $\lambda \mathbf{x}^* + (1-\lambda)\mathbf{x}^* = \mathbf{x}^* \in \Gamma$.

Finalmente podemos suponer que Γ tiene al menos dos elementos $\mathbf{x}^*$, $\mathbf{y}^* \in \Gamma$. Por la definición de Γ, entonces

$$f_{\min} = \min_{x \in \Omega} f(x) = f(\mathbf{x}^*) = f(\mathbf{y}^*)$$

Si tomamos $\lambda \in [0,1]$, por convexidad de $f(\mathbf{x})$ se cumple

$$f(\lambda \mathbf{x}^* + (1-\lambda)\mathbf{y}^*) \leq \lambda f(\mathbf{x}^*) + (1-\lambda) f(\mathbf{y}^*) = \lambda f_{\min} + (1-\lambda) f_{\min} = f_{\min}$$

pero como $\lambda \mathbf{x}^* + (1-\lambda)\mathbf{y}^* \in \Omega$ y $f_{\min} = \min_{x \in \Omega} f(x)$, tendrá que ocurrir

$$f_{\min} \leq f(\lambda \mathbf{x}^* + (1-\lambda)\mathbf{y}^*)$$

de donde

$$f_{\min} \leq f(\lambda \mathbf{x}^* + (1-\lambda)\mathbf{y}^*) \leq f_{\min}$$

y por tanto

$$f(\lambda \mathbf{x}^* + (1-\lambda)\mathbf{y}^*) = f_{\min}$$

y se obtiene el resultado pedido.

Supongamos ahora que $\mathbf{x}^* \in \Omega$ es un mínimo local de $f(\mathbf{x})$ que no es global; es decir, que $\exists\ \mathbf{y} \in \Omega$ con $f(\mathbf{y}) < f(\mathbf{x}^*)$. Por la convexidad de $f(\mathbf{x})$ en el conjunto convexo Ω se obtiene la siguiente desigualdad

$$f(\lambda \mathbf{y} + (1-\lambda)\mathbf{x}^*) \leq \lambda f(\mathbf{y}) + (1-\lambda) f(\mathbf{x}^*)$$

para $\lambda \in [0,1]$. Utilizando ahora el hecho de que $f(\mathbf{y}) < f(\mathbf{x}^*)$ y como $\lambda \in [0,1]$ y por tanto tanto λ como $(1-\lambda)$ son no negativos

$$f(\lambda \mathbf{y} + (1-\lambda)\mathbf{x}^*) \leq \lambda f(\mathbf{x}^*) + (1-\lambda) f(\mathbf{x}^*) = f(\mathbf{x}^*)$$

Para un valor de λ suficientemente pequeño se contradice el hecho de que $\mathbf{x}^*$ sea un mínimo relativo, como estamos asuminedo que $\mathbf{x}^*$ es mínimo local, entonces debe cumplirse que además es global. $\blacksquare$.

Teorema 1.9 *Sea $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ con Ω un conjunto convexo con $f \in \mathcal{C}^1(\Omega)$ una función convexa [cóncava]. Supongamos que existe un $\mathbf{x}^* \in \Omega$ que cumple la siguiente propiedad*

$$\forall \mathbf{y} \in \Omega \Rightarrow \nabla f(\mathbf{x}^*)^T (\mathbf{y} - \mathbf{x}^*) \geq 0 \quad \left[\nabla f(\mathbf{x}^*)^T (\mathbf{y} - \mathbf{x}^*) \leq 0\right]$$

Entonces $\mathbf{x}^*$ es un punto de mínimo [máximo] global de f en Ω.

Demostración: Como Ω convexo, $\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*) \in \Omega$ con $\lambda \in [0,1]$. Por la caracterización de primer orden de funciones convexas (proposición 1.6) y utilizando el hecho de que para $\mathbf{x}^*$ ocurre $\nabla f(\mathbf{x}^*)^T(\mathbf{y} - \mathbf{x}^*) \geq 0$ tenemos:

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T(\mathbf{y} - \mathbf{x}^*) \geq f(\mathbf{x}^*)$$

Luego $\mathbf{x}^*$ es un punto de mínimo relativo de f en Ω y utilizando el resultado de la proposición anterior, también es un mínimo global $\blacksquare$.

Se incluyen a continuación, sin demostración, una serie de resultados muy útiles e interesantes para la resolución de problemas de optimización.

Proposición 1.10 *Si Ω es un subconjunto no vacío, convexo y compacto de $\mathbb{R}^n$ cuyo conjunto de puntos extremos es finito y si $f : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ es un función convexa [cóncava] sobre $\Omega \Rightarrow f(\mathbf{x})$ posee en Ω un máximo [mínimo] global y se encuentra en uno de sus puntos extremos.*

Teorema 1.11 *Sea $f : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ convexa [cóncava] en Ω convexo y compacto $\Longrightarrow$ Si $f(\mathbf{x})$ tiene máximo [mínimo] en Ω entonces lo alcanza en un punto extremo de Ω.*

Lema 1.12 *Si $\mathbf{x}^*, \mathbf{y}^*$ son dos puntos de mínimo [máximo] global de una función convexa [cóncava] $f(\mathbf{x})$, $f : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ con Ω convexo $\Longrightarrow \forall\lambda \in [0,1]$ los puntos definidos como $\mathbf{z}^*(\lambda) = \lambda\mathbf{x}^* + (1-\lambda)\mathbf{y}^*$ también son mínimos [máximos] globales de $f(\mathbf{x})$.*

Demostración: Supongamos que f es convexa (para el caso de f cóncava y máximo la demostración sería equivalente), entonces

$$f(\mathbf{z}^*(\lambda)) = f(\lambda\mathbf{x}^* + (1-\lambda)\mathbf{y}^*) = f(\lambda\mathbf{x}^* + (1-\lambda)\mathbf{y}^*) \leq \lambda f(\mathbf{x}^*) + (1-\lambda) f(\mathbf{y}^*)$$

Como $\mathbf{x}^*, \mathbf{y}^*$ son mínimos globales $\Longrightarrow f(\mathbf{y}^*) = f(\mathbf{x}^*)$ y resulta

$$f(\mathbf{z}^*(\lambda)) \leq f(\mathbf{x}^*)$$

pero como $\mathbf{x}^*$ es el mínimo global de $f(\mathbf{x})$ sobre Ω, que es un conjunto convexo, ocurre

$$\mathbf{z}^*(\lambda) \in \Omega \Rightarrow f(\mathbf{z}^*(\lambda)) = f(\mathbf{x}^*)$$

y los puntos de la forma $\mathbf{z}^*(\lambda)$, con $\lambda \in [0,1]$, son todos mínimos globales de f en Ω. $\blacksquare$

Lema 1.13 *Sea $f : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, con Ω un conjunto convexo no vacío y $f(\mathbf{x})$ convexa [cóncava]. Sea $\mathbf{x}^*$ un mínimo [máximo] local de f. Entonces, si $\mathbf{x}^*$ es un mínimo [máximo] local estricto de $f(\mathbf{x})$ o si $f(\mathbf{x})$ es estrictamente convexa [cóncava] sobre Ω, entonces $\mathbf{x}^*$ es el único mínimo [máximo] global de $f(\mathbf{x})$ en Ω.*

Demostración: Si $\mathbf{x}^*$ es un mínimo local entonces por el teorema 1.8 es un mínimo global. Si suponemos ahora que existe otro punto $\mathbf{z}^*$ mínimo de $f(\mathbf{x})$, por el lema 1.12 el mínimo también se alcanza en todo el segmento que une ambos puntos $\mathbf{x}^*$ y $\mathbf{z}^*$ y esto contradice el hecho de que $\mathbf{x}^*$ sea estricto.

Supongamos ahora que $\mathbf{x}^*$ es un mínimo local de $f(\mathbf{x})$ y que $f(\mathbf{x})$ es estrictamente convexa. Como $f(\mathbf{x})$ es convexa, ya que es estrictamente convexa, $\mathbf{x}^*$ es un mínimo global. Si ahora suponemos que existe otro mínimo global $\mathbf{z}^*$, por la convexidad estricta de f obtenemos:

$$f(\frac{1}{2}\mathbf{x}^* + \frac{1}{2}\mathbf{z}^*) < \frac{1}{2}f(\mathbf{x}^*) + \frac{1}{2}f(\mathbf{z}^*) = f(\mathbf{x}^*)$$

y por tanto $\mathbf{x}^*$ no sería mínimo global $\blacksquare$.

Problemas propuestos

Ejercicio 1.1 *Demuestra que los hiperplanos de* $\mathbb{R}^n$ *son conjuntos convexos.*

Ejercicio 1.2 *Demuestra que todos los semiespacios (positivos y/o negativos) asociados a un hiperplano* $H \subseteq \mathbb{R}^n$ *son conjuntos convexos.*

Ejercicio 1.3 *Demuestra el lema 1.2. Si* Ω_1 *y* Ω_2 *son dos subconjuntos convexos no vacíos de* $\mathbb{R}^n$ *entonces los siguientes conjuntos son convexos:*

a) $\Omega_1 + \Omega = \{\mathbf{x}_1 + \mathbf{x}_2 \in \mathbb{R}^n : \mathbf{x}_1 \in \Omega_1; \mathbf{x}_2 \in \Omega_2\}$

b) $\Omega_1 - \Omega = \{\mathbf{x}_1 - \mathbf{x}_2 \in \mathbb{R}^n : \mathbf{x}_1 \in \Omega_1; \mathbf{x}_2 \in \Omega_2\}$

c) $\Omega_1 \cap \Omega_2$

Ejercicio 1.4 *Prueba la convexidad del conjunto* Ω *definido por*

$$\Omega = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0; i = 1, \ldots, n \right\}$$

Ejercicio 1.5 *Demuestra la convexidad del conjunto*

$$\Omega = \left\{(x, y) \in \mathbb{R}^2 : y \leq 3 - x^2\right\}$$

Ejercicio 1.6 *Comprueba la convexidad del conjunto*

$$\Omega = \left\{(x, y) \in \mathbb{R}^2 : y = e^x\right\}$$

Ejercicio 1.7 *Siendo* $r > 0$ *y* $\mathbf{a} \in \mathbb{R}^n$*, prueba que la bola abierta de centro* $\mathbf{a}$ *y radio* r *es un conjunto convexo.*

Ejercicio 1.8 *Demuestra que* $\mathbf{z} \in \mathbb{R}^n$ *es combinación lineal convexa de* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *si y sólo si* $\mathbf{z} \in [\mathbf{x}, \mathbf{y}]$.

Ejercicio 1.9 *Prueba que los siguientes conjuntos son convexos*

a) $\Omega_1 = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 2,\ x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0\right\}$

b) $\Omega_2 = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid 3x_1 + 2x_2 \leq 4,\ 6x_1 + x_2 \leq 3,\ x_1 \geq 0,\ x_2 \geq 0\right\}$

c) $\Omega_3 = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid \frac{x_1^2}{2} + x_2^2 \leq 1,\ x_2 \geq 0\right\}$

d) $\Omega_4 = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 \leq 3,\ x_1 + x_2 + x_3 \geq 1\right\}$

Ejercicio 1.10 *Estudia la concavidad o convexidad en* $\mathbb{R}^2$ *o* $\mathbb{R}^3$ *de las siguientes funciones*

a) $f_1(x_1, x_2) = (x_1 - 3)^2 + (x_2 + 1)^2$

b) $f_2(x_1, x_2) = (x_1 - 2)^2 + x_2^3$

c) $f_3(x_1, x_2) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & (x_1, x_2) \neq (0, 0) \\ 0 & (x_1, x_2) = (0, 0) \end{cases}$

d) $f_4(x_1, x_2, x_3) = x_1^2 - x_2^2 + x_3^2 + x_1 x_2$

Ejercicio 1.11 *¿Será convexa la función $f_2(x_1, x_2)$ del problema anterior sobre el conjunto abierto y convexo*

$$\Omega_1 = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 > 0,\ x_2 > 0\right\}?$$

¿Y $f_4(x_1, x_2)$ sobre el conjunto abierto convexo

$$\Omega_2 = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 > 0,\ x_2 > 0,\ x_3 > 0\right\}?$$

Ejercicio 1.12 *Para los problemas de optimización siguientes*

$$\begin{array}{lc} \text{(a) Optimizar} & x_1 + x_2 \\ \text{sujeto a} & 9x_1 + 2x_2 \leq 6 \\ & 2x_1 + 2x_2 \leq 3 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{array} \qquad \begin{array}{lc} \text{(b) Optimizar} & (x_1 - 1)^2 + (x_2 - 1)^2 \\ \text{sujeto a} & x_1 + 2x_2 \leq 4 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{array}$$

se pide la resolución razonada de los siguientes apartados:

1. *El análisis de su convexidad.*
2. *¿Qué se puede decir de sus extremos locales y globales?*
3. *La resolución geométrica del problema.*

Ejercicio 1.13 *Dada la función*

$$f(x) = x^p \quad 0 < p \leq 1$$

Responde de forma razonada a cada uno de los siguientes apartados

1. *Estudia la concavidad o convexidad de la función $f(x)$ sobre el intervalo $I = (0, \infty)$.*
2. *Utilizando la información del apartado anterior, demuestra la siguiente desigualdad*

$$(x + y)^p \geq 2^{p-1}(x^p + y^p) \qquad x, y > 0$$

Ejercicio 1.14 *Sea P_1 un problema de programación matemática y sea P_2 otro problema que resulta de anadirle a P_1 una restricción más. Supongamos que el objetivo es maximizar. Si Ω_1 y Ω_2 son respectivamente sus conjuntos factibles, resuelve los siguientes apartados:*

1. *Indica si existe alguna relación entre los conjuntos Ω_1y Ω_2.*
2. *Supongamos que $x^* \in \Omega_1$ es solución óptima de P_1 ¿será x^* solución óptima de P_2?*
3. *Suponiendo que ambos problemas tienen soluciones óptimas x_1^* y x_2^* respectivamente. Encuentra, si existe, la relación entre ellas.*