Capítulo 1

Fundamentos de Optimización Estática

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 un volumen máximo, pero utilizando para ello una cantidad fija de material. El problema en forma abstracta se podría plantear en los siguientes términos*

$$\begin{array}{ll} \text{Maximizar} & \text{Volumen} \\ \text{s.a.} & \text{Área Fija} \end{array}$$

Para resolver este problema, en primer lugar, tendremos que modelizarlo matemáticamente, es decir tendremos que expresar el problema en términos matemáticos. La forma de modelizar un problema de optimización es empezar por definir las variables que están implicadas en dicho problema y puesto que estamos tratando de determinar las dimensiones de una caja rectangular, la opción más clara es considerar como variables a 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 fácilmente como

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

A continuación tenemos que tener en cuenta las limitaciones de material existentes. Como este material se utiliza para construir las paredes de la caja, tendremos que tener en cuenta el área lateral de la misma, que en el caso de utilizar tapa para la caja sería

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

Por último, es necesario tener en cuenta que las dimensiones de la caja no pueden ser negativas y entonces el problema puede plantearse como

$$\begin{array}{lc} \text{Maximizar} & xyz \\ \text{s.a.} & 2(xy + yz + zx) = A \\ & x, y, z \geq 0 \end{array}$$

A partir de este sencillo ejemplo se observa que cada problema de optimización tiene tres elementos fundamentales:

1. *Variables de decisión:* El primer elemento clave en la formulación de problemas de optimización es la selección de las variables independientes que son adecuadas para caracterizar los posibles diseños candidatos o las condiciones de funcionamiento del sistema. Hay diversos factores a considerar en la elección de las variables independientes aunque una buena regla es incluir como variables aquellas que tengan un impacto significante sobre la función objetivo.

   En esta parte de la asignatura, representaremos las variables independientes, mediante vectores columna de $\mathbb{R}^n$, en la forma
$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$
   Aunque para los casos $n = 1$, 2 y 3, emplearemos la notación usual de x, (x, y) y (x, y, z) respectivamente.

2. *Restricciones:* Una vez determinadas las variables independientes, el siguiente paso en la formulación de problemas de optimización es establecer, mediante ecuaciones o inecuaciones las relaciones existentes entre las variables de decisión. Estas relaciones son debidas a limitaciones en el sistema, a leyes naturales o a limitaciones tecnológicas, entre otras. El conjunto de restricciones representa las limitaciones del sistema.

   Podemos distinguir dos tipos de relaciones:

   (a) *Restricciones de igualdad:* Son relaciones entre las variables de la forma
$$h(x_1, \ldots, x_n) = 0$$
   donde $h : A \subseteq \mathbb{R} \to \mathbb{R}$, es una función real de variables reales.

   (b) *Restricciones de desigualdad:* Son relaciones entre las variables de la forma
$$g(x_1, \ldots, x_n) \leq 0$$
   donde $g : A \subseteq \mathbb{R} \to \mathbb{R}$, es una función real de variables reales.

   Observación 1.2 *Solamente se han considerado restricciones de dos tipos, $h(x) = 0$ y $g(x) \leq 0$, puesto que siempre es posible, mediante una simple transformación, expresar el problema en términos de este tipo de restricciones, tal y como se puede apreciar en la siguiente tabla:*
$$\begin{array}{ccccc} h(x_1, \ldots, x_n) = b & \Leftrightarrow & \widehat{h} = h - b & \Leftrightarrow & \widehat{h}(x_1, \ldots, x_n) = 0 \\ g(x_1, \ldots, x_n) \leq c & \Leftrightarrow & \widehat{g} = g - c & \Leftrightarrow & \widehat{g}(x_1, \ldots, x_n) \leq 0 \\ g(x_1, \ldots, x_n) \geq c & \Leftrightarrow & \widehat{g} = c - g & \Leftrightarrow & \widehat{g}(x_1, \ldots, x_n) \leq 0 \end{array}$$

   Observación 1.3 *Normalmente, dentro de este conjunto de restricciones, no se incluyen las condiciones sobre las variables del tipo $x \geq 0$ o similares.*

3. *Función objetivo:* También llamado índice de rendimiento o criterio de elección, el último elemento de un problema de optimización es la elección adecuada del criterio sobre el que el desarrollo o diseño del sistema puede evaluarse de manera que podamos identificar el “mejor” diseño para ese criterio. Por ejemplo, en muchas aplicaciones de ingeniería se suele elegir un criterio económico, sin embargo, hay muchas opciones que permiten considerar dicho criterio: coste del capital total, coste anual, beneficio anual neto, tasa coste-beneficio. Para otras aplicaciones el criterio de elección puede implicar factores tecnológicos, por ejemplo, tiempo de producción mínimo, máxima tasa de producción, mínima energía utilizada o máxima carga, entre otros. Independientemente del criterio seleccionado, en el contexto de optimización el “mejor” siempre indica el candidato del sistema que produce el mínimo o máximo valor, según el criterio utilizado, para la función objetivo elegida.

   Es importante hacer notar que en este libro se utilizará un único criterio de optimización para definir el óptimo. 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 es una simplificación importante, puesto que 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 criterios como secundarios. El criterio primario se utiliza como medida de optimización del proceso, 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*.

Después de introducir los elementos básicos de un problema de optimización, nuestro objetivo está más clarificado: A partir del valor de la *función objetivo*, diseñada para cuantificar el rendimiento y medir la calidad de la decisión, se obtienen valores para un cierto número de *variables de decisión*, de forma que éstas minimicen o maximicen dicha función objetivo. Por lo general, estas variables estarán relacionadas entre sí mediante expresiones matemáticas o *restricciones* que limitan la elección de esos valores. De otra forma, un problema de optimización consiste en buscar valores para determinadas variables de forma que, cumpliendo un conjunto de requisitos representados en general mediante ecuaciones y/o inecuaciones algebráicas, proporcionen el mejor valor posible para una función que es utilizada para medir el rendimiento del sistema en estudio. Con el mejor valor queremos indicar el mayor o el menor valor posible, según la naturaleza del problema. Buscamos, en resumen, valores que cumplan unas condiciones y minimicen o maximicen una función que caracteriza el sistema. Con esta idea, el planteamiento en abstracto general para resolver problemas de este tipo es el siguiente

$$\begin{array}{cc} \text{Optimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & \textit{Restricciones} \end{array}$$

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

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

A continuación se incluyen dos ejemplos con esta estructura.

Ejemplo 1.4 ***Distancia más corta entre dos curvas:*** *Supongamos que queremos 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 cortan 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*

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

o más concretamente como

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

siendo

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

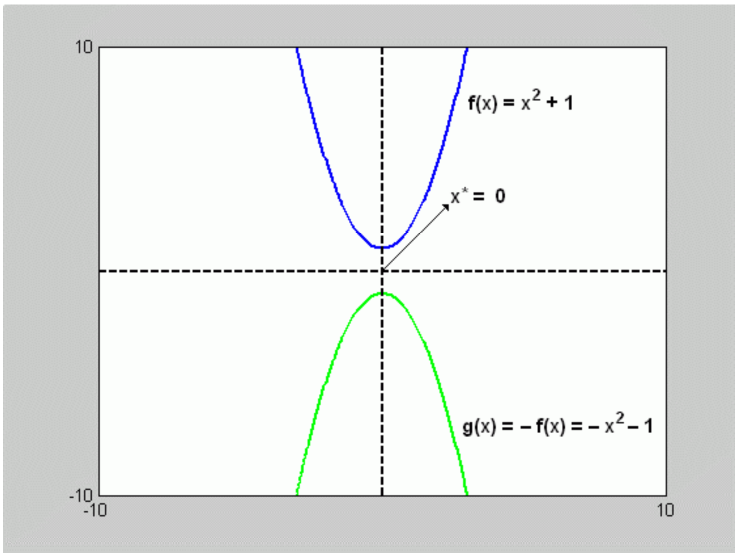

Figura 1.1: Equivalencia entre $\min f(x)$ **y** $\max g(x) = -f(x)$**, en ambos casos es** $x^* = 0$**.**

las coordenadas de los dos puntos.

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

Ejemplo 1.5 ***Problema lineal.*** *Si 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:*

$$\begin{array}{cc} \textit{Minimizar} & b_1x_1 + b_2x_2 \\ \textit{Sujeto a} & c_1x_1 + c_2x_2 \leq P \\ & x_1 \geq 0; x_2 \geq 0 \end{array}$$

donde P *es el presupuesto total disponible y los parámetros* b_j *y* c_j $j = 1, 2$ *son el beneficio y el coste, respectivamente, para cada uno de los productos.*

1.2 Definiciones

Daremos a continuación una serie de definiciones elementales relacionadas con la teoría de la optimización matemática con el objetivo de que el lector se familiarice con este tipo de problemas.

Definición 1.6 *Definiremos el problema fundamental de la optimización estática o problema de programación no lineal (*Nonlinear Programming Problem o NLPP*) al expresado en los siguientes términos:*

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

En principio las funciones implicadas en la definición del problema fundamental no tienen porqué tener ninguna propiedad 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, g_j son continuas y que en la mayoría de los casos tienen derivadas primeras y segundas también continuas. Además el conjunto A será en la mayoría de los casos un conjunto convexo.

La resolución del problema de optimización 1.1 consistirá en primer lugar, en buscar valores para las variables x_i que resuelvan el sistema formado por las restricciones y en segundo lugar encontrar de entre estos valores aquel o aquellos que proporcionen el mayor (maximizar) o menor (minimizar) valor para la función real $f(x_1, \ldots, x_n)$.

Definición 1.7 *Casos especiales del problema NLPP:*

1. Problema sin restricciones: *En este caso no hay restricciones de ningún tipo, es decir, $m = p = 0$. La expresión general de estos problemas sería la siguiente:*

$$\begin{array}{rl} \textit{Optimizar} & f(x_1, \ldots, x_n) \\ \textit{Sujeto a} & (x_1, \ldots, x_n) \in A \end{array}$$

Las únicas limitaciones vienen dadas por el conjunto A donde esté definida la función f.

2. Problema de Lagrange o con restricciones de igualdad: *En este caso, en el planteamiento del problema solamente existen restricciones de igualdad, es decir, $m \neq 0$ y $p = 0$:*

$$\begin{array}{rl} \textit{Optimizar} & f(x_1, \ldots, x_n) \\ \textit{Sujeto a} & h_i(x_1, \ldots, x_n) = 0 \quad i = 1, \ldots, m \\ & (x_1, \ldots, x_n) \in A \end{array}$$

3. Problemas unidimensionales o univariantes: *Este es un caso particular de los problemas sin restricciones en los que solamente hay una variable, es decir, cuando $n = 1$, $m \neq 0$ y $p \neq 0$. La estructura sería*

$$\begin{array}{rl} \textit{Optimizar} & f(x) \\ \textit{Sujeto a} & x \in I \subseteq \mathbb{R} \end{array}$$

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

Definición 1.8 (solución factible) *Diremos que $\overline{\mathbf{x}} = (\overline{x}_1, \ldots, \overline{x}_n) \in A \subseteq \mathbb{R}$ es una* solución factible *del problema 1.1 si cumple todas sus restricciones, es decir, si*

$$\begin{array}{rcl} h_i(\overline{x}_1, \ldots, \overline{x}_n) & = & 0 \quad i = 1, \ldots, m \\ \\ g_j(\overline{x}_1, \ldots, \overline{x}_n) & \leq & 0 \quad j = 1, \ldots, p \end{array}$$

Definición 1.9 (conjunto factible) *Se define* región *o* conjunto factible *del problema* NLPP, *que denotaremos en general por Ω, al conjunto de todas sus soluciones factibles*

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

Nuestro objetivo al intentar resolver el problema de optimización 1.1 es encontrar la "mejor" de las soluciones factibles.

Definición 1.10 *Dado el problema* NLPP *de programación no lineal descrito en la ecuación 1.1 y sea Ω su conjunto factible. Diremos que $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ es un* punto de mínimo global *para el problema si*

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

El punto $\mathbf{x}^*$ será mínimo global estricto *si la desigualdad es estricta es decir si $\forall\, \overline{\mathbf{x}} \in \Omega$, con $\overline{\mathbf{x}} \neq \mathbf{x}^*$, ocurre $f(\mathbf{x}^*) < f(\overline{\mathbf{x}})$.*

Definición 1.11 *Dado el problema* NLPP *de programación no lineal descrito en la ecuación 1.1 y sea Ω su conjunto factible. Diremos que $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ es un* punto de máximo global *para el problema si*

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

El punto $\mathbf{x}^*$ será máximo global estricto *si la desigualdad es estricta es decir si $\forall\, \overline{\mathbf{x}} \in \Omega$, con $\overline{\mathbf{x}} \neq \mathbf{x}^*$, ocurre $f(\mathbf{x}^*) > f(\overline{\mathbf{x}})$.*

Observación 1.12 *Los máximos y mínimos globales de un problema de programación no lineal también se denominan* extremos globales.

Definición 1.13 *Dado el problema* NLPP *de programación no lineal descrito en la ecuación 1.1 y sea* Ω *su conjunto factible. Diremos que* $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ *es una* solución óptima *del problema si*

1. *Es un mínimo global y el objetivo del problema es minimizar.*
2. *Es un máximo global y el objetivo del problema es maximizar.*

Resolver el 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 y que definiremos a continuación.

Definición 1.14 *Dado el problema* NLPP *de programación no lineal descrito en la ecuación 1.1 y sea* Ω *su conjunto factible. Diremos que* $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ *es un* punto de mínimo local o relativo *para el problema 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 de* mínimo local estricto *si la desigualdad es estricta es decir si ocurre* $f(\mathbf{x}^*) < f(\overline{\mathbf{x}})$ *cuando* $\|\overline{\mathbf{x}} - \mathbf{x}^*\| < \epsilon$.

Definición 1.15 *Dado el problema* NLPP *de programación no lineal descrito en la ecuación 1.1 y sea* Ω *su conjunto factible. Diremos que* $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ *es un* punto de máximo local o relativo *para el problema 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 de* máximo local estricto *si la desigualdad es estricta es decir si ocurre* $f(\mathbf{x}^*) > f(\overline{\mathbf{x}})$ *cuando* $\|\overline{\mathbf{x}} - \mathbf{x}^*\| < \epsilon$.

Definición 1.16 *Los máximos y mínimos locales de un problema de programación no lineal también se denominan* extremos locales o relativos.

Observación 1.17 *Los extremos locales o globales estrictos, también reciben el nombre de* extremos fuertes, *mientras que para los no estrictos se suele emplear el término débil.*

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 deben cumplir las funciones $f(\mathbf{x})$, $h_i(\mathbf{x})$ y $g_j(\mathbf{x})$ para que un problema tenga solución?
2. **Cuestión dinámica:** Si $\mathbf{x}^*$ no es el punto óptimo, ¿cómo podemos encontrar una solución que sea óptima utilizando la información de la función en $\mathbf{x}^*$?

Finalmente, un problema de optimización o programación matemática puede tener o no solución, e incluso si hay solución al problema, puede que no sea única. El resultado principal utilizado para conocer si un problema de optimización tiene solución es el *teorema de Weierstrass*.

Teorema 1.18 (Teorema de Weierstrass) *Sea* $f(\mathbf{x})$ *una función continua definida sobre un conjunto compacto (cerrado y acotado)* $K \subseteq \mathbb{R}^n$*, entonces siempre existe solución al problema (tanto de minimización como de maximización). Es decir*

$$\exists \quad \mathbf{x}^{\min}, \mathbf{x}^{\max} \in K \Rightarrow \begin{array}{l} f\left(\mathbf{x}^{\min}\right) = \min\limits_{\mathbf{x}\in K} f(\mathbf{x}) \\ \\ f\left(\mathbf{x}^{\max}\right) = \max\limits_{\mathbf{x}\in K} f(\mathbf{x}) \end{array}$$

Este es un resultado importante a tener en cuenta para la existencia de puntos óptimos, sin embargo no nos proporciona un método para la localización de los mismos, 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.

1.3 Conjuntos Convexos

El concepto de convexidad es de gran importancia en el estudio de los problemas de optimización. La noción de convexidad es interesante desde el punto de vista de la aplicación práctica, puesto que en algunos casos, por ejemplo en los llamados problemas convexos, se puede garantizar que un óptimo local de un problema es realmente un óptimo global del mismo.

Se describen en esta sección algunos conceptos indispensables para el desarrollo de la programación matemática. Aunque se puede definir convexidad en el ámbito de cualquier espacio topológico, se considerará el espacio vectorial $\mathbb{R}^n$.

Definición 1.19 (segmento lineal) *Dados dos puntos* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*, el* segmento lineal cerrado *que los une es el conjunto*

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

de igual forma, el segmento lineal abierto *que une* $\mathbf{x}$ *e* $\mathbf{y}$ *es el conjunto*

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

Definición 1.20 (conjunto convexo) *Se dice que un conjunto* $\Omega \subset \mathbb{R}^n$ *es* convexo *si verifica:*

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

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

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

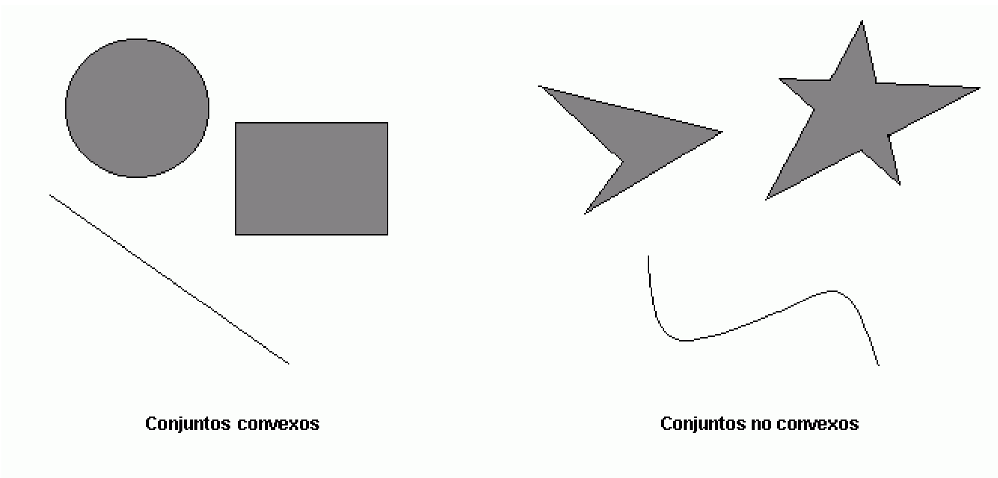

Figura 1.2: Convexidad en $\mathbb{R}^2$.

Uno de los tipos más importantes de conjunto convexo es el *hiperplano.*

Definición 1.21 (hiperplano) *Si* $\mathbf{a} \in \mathbb{R}^n; \mathbf{a} \neq 0, c \in \mathbb{R}$*. El conjunto*

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

es un hiperplano *de* $\mathbb{R}^n$*. El vector* $\mathbf{a}$ *es llamado* normal *al hiperplano. Se puede comprobar que* H *es un conjunto convexo.*

Definición 1.22 (semiespacios) *Sea* $\mathbf{a} \in \mathbb{R}^n$ *con* $\mathbf{a} \neq \mathbf{0}$ *y* $b \in \mathbb{R}$. *Sea* H *el hiperplano construido con* $\mathbf{a}$ *y* b, *entonces definimos* semiespacios cerrados positivos *y* negativos, *respectivamente a los conjuntos siguientes:*

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

y semiespacios abiertos positivos y negativos a:

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

que son todos conjuntos convexos.

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 se deja como ejercicio.

Lema 1.23 *Sean* Ω_1 *y* Ω_2 *conjuntos convexos de* $\mathbb{R}^n$. *Entonces:*

1. $\Omega_1 \cap \Omega_2$ *es convexo.*
2. $\Omega_1 + \Omega_2 = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in \Omega_1, \mathbf{y} \in \Omega_2\}$ *es convexo.*
3. $\Omega_1 - \Omega_2 = \{\mathbf{x} - \mathbf{y} : \mathbf{x} \in \Omega_1, \mathbf{y} \in \Omega_2\}$ *es convexo.*

Teniendo en cuenta que los semiespacios de $\mathbb{R}^n$ son conjuntos convexos y también los resultados del lema anterior se obtiene que los conjuntos de la forma

$$P = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$$

donde $\mathbf{A}$ es una matriz $m \times n$ en $\mathbb{R}$ y $\mathbf{b} \in \mathbb{R}^m$ vector son conjuntos convexos puesto que son intersecciónes de m semiespacios.

El conjunto $S = \{\mathbf{x} \mid \mathbf{x} \text{ es solución de } P\}$, siendo P el problema siguiente:

$$P = \begin{cases} \text{Minimizar} & \mathbf{c}^T\mathbf{x} \\ \text{Sujeto a} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{cases}$$

también es un conjunto convexo.

La siguiente proposición indica que las aplicaciones lineales conservan la convexidad.

Proposición 1.24 *Sea* $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, *con* Ω *convexo y* $f(\mathbf{x})$ *lineal* $\Rightarrow$ *El conjunto imagen* $f(\Omega)$ *es un conjunto convexo.*

Demostración: Si $f(\mathbf{x})$ es una aplicación lineal entonces cumple

$$f(\lambda\mathbf{x}+\mu\mathbf{y}) = \lambda f(\mathbf{x}) + \mu f(\mathbf{y})$$

cualesquiera que sean λ y μ. Hay que comprobar que $f(\Omega)$ es un conjunto convexo.

$$f(\Omega) = \{f(\mathbf{x}) \mid \quad \mathbf{x} \in \Omega\}$$

Para ello tomamos dos elementos $\mathbf{y}^1$, $\mathbf{y}^2 \in f(\Omega)$, es decir

$$\exists\ \mathbf{x}^1 \in \Omega \text{ tal que } \mathbf{y}^1 = f\left(\mathbf{x}^1\right)$$

$$\exists\ \mathbf{x}^2 \in \Omega \text{ tal que } \mathbf{y}^2 = f\left(\mathbf{x}^2\right)$$

Como Ω es un conjunto convexo debe ocurrir

$$\mathbf{x}^1, \mathbf{x}^2 \in \Omega \Rightarrow \lambda \mathbf{x}^1 + (1-\lambda)\,\mathbf{x}^2 \in \Omega \qquad \forall \lambda \in [0,1]$$

de donde

$$f\left(\lambda \mathbf{x}^1 + (1-\lambda)\,\mathbf{x}^2\right) \in f(\Omega)$$

teniendo en cuenta la propiedad de linealidad de $f(\mathbf{x})$

$$f\left(\lambda \mathbf{x}^1 + (1-\lambda)\,\mathbf{x}^2\right) = \lambda f\left(\mathbf{x}^1\right) + (1-\lambda)\, f\left(\mathbf{x}^2\right) = \lambda \mathbf{y}^1 + (1-\lambda)\,\mathbf{y}^2 \in f(\Omega) \qquad \forall \lambda \in [0,1]$$

y por tanto $f(\Omega)$ es convexo.

La siguiente definición extiende el concepto de segmento lineal (cerrado o abierto) formado por combinaciones lineales de dos puntos al caso de más de dos puntos.

Definición 1.25 (combinación lineal convexa) *Sea* $\mathbf{x} \in \mathbb{R}^n$*, diremos que es* combinación lineal convexa *de los puntos* $\mathbf{x}^1, \ldots, \mathbf{x}^m \in \mathbb{R}^n$ *si existen* m *números reales* $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$*, que cumplen*

$$\lambda_j \geq 0 \quad j = 1, \ldots, m$$

$$\sum_{j=1}^{m} \lambda_j = 1$$

$$\mathbf{x} = \sum_{j=1}^{m} \lambda_j \mathbf{x}^j$$

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

Definición 1.26 (envoltura convexa) *Sea* $\Omega \subseteq \mathbb{R}^n$*. Se define la* envoltura convexa *de* Ω*, y la denotamos por* $H(\Omega)$*, al conjunto de todas las combinaciones lineales convexas de* Ω *formadas por un número finito de elementos de* Ω*, es decir*

$$\mathbf{x} \in H(\Omega) \Longleftrightarrow \mathbf{x} = \textstyle\sum_{j=1}^{m} \lambda_j \mathbf{x}^j$$

$$\textstyle\sum_{j=1}^{m} \lambda_j = 1$$

$$\lambda_j \geq 0, j = 1, \ldots, m$$

donde m *es un entero positivo y* $\{\mathbf{x}^1, \ldots, \mathbf{x}^m\} \in \Omega$*. Está claro que* $\Omega \subseteq H(\Omega)$*, puesto que si* $\mathbf{x} \in \Omega \Rightarrow \mathbf{x} = 1 * \mathbf{x}$ *es una combinación lineal convexa, luego* $\mathbf{x} \in \Omega$*. Por construcción,* $H(\Omega)$ *es un conjunto convexo. Es el menor conjunto convexo que contiene a* Ω*.*

Por definición, un punto de la envoltura convexa de un conjunto se puede representar como combinación convexa de un número finito de puntos del conjunto.

Podemos observar en la figura 1.3 un conjunto no convexo y su envoltura convexa.

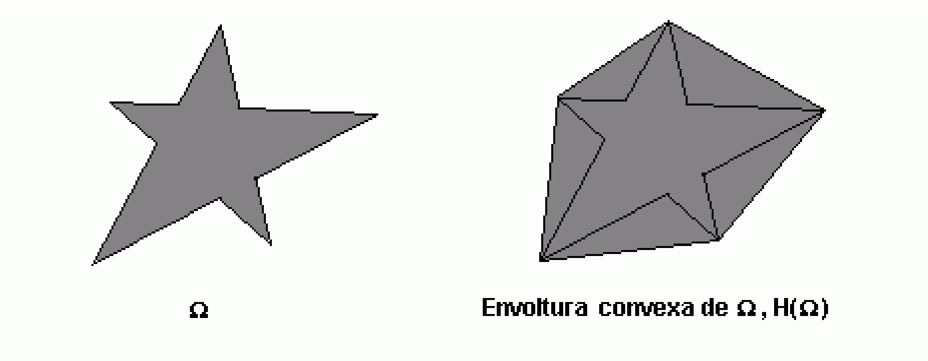

Figura 1.3: Envoltura convexa de un conjunto.

Definición 1.27 (polítopo, simplex, poliedro) *La envoltura convexa de un conjunto finito de puntos de* $\mathbb{R}^n$*,* $\Omega = \{\mathbf{x}^1, \ldots, \mathbf{x}^{k+1}\}$*, se denomina* polítopo*. Si además el conjunto* $\{\mathbf{x}^2 - \mathbf{x}^1, \mathbf{x}^3 - \mathbf{x}^1, \ldots, \mathbf{x}^{k+1} - \mathbf{x}^1\}$ *es linealmente independiente, entonces esta envoltura convexa* $H(\Omega) = H(\mathbf{x}^1, \ldots, \mathbf{x}^{k+1})$ *se llama* simplex *de vértices* $\mathbf{x}^1, \ldots, \mathbf{x}^{k+1}$*. Por último un polítopo acotado se denomina* poliedro.

Es posible demostrar que los polítopos son la intersección de un número finito de semiespacios cerrados y equivalentemente a un conjunto de desigualdades lineales de la forma

$$\left.\begin{array}{c} \mathbf{a}_1^T\mathbf{x} \leq b_1 \\ \mathbf{a}_2^T\mathbf{x} \leq b_2 \\ \vdots \\ \mathbf{a}_n^T\mathbf{x} \leq b_n \end{array}\right\} \Leftrightarrow \mathbf{A}\mathbf{x} \leq \mathbf{b}$$

Definición 1.28 (punto extremo) *Dado un conjunto* $\Omega \subseteq \mathbb{R}^n$*, un conjunto convexo no vacío,* $\mathbf{x} \in \Omega$*. El punto* $\mathbf{x}$ *es un* vértice o punto extremo *de* $\Omega \Longleftrightarrow$

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

Es decir un punto extremo es un punto del conjunto que no puede expresarse como punto medio de ningún par de puntos del conjunto.

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 1.4).

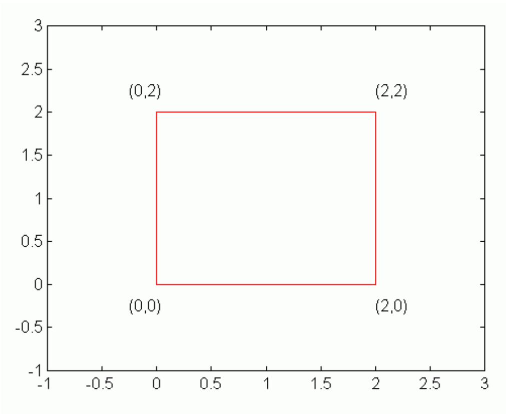

Figura 1.4: Puntos extremos.

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 1.4 o tener una cantidad infinita de elementos, como una bola cerrada de $\mathbb{R}^n$.

1.4 Funciones convexas

Las funciones convexas tienen muchas propiedades especiales e importante. Por ejemplo, veremos que cualquier mínimo local de una función convexa sobre un conjunto convexo es también un mínimo global. En esta sección se proporciona la definición de función convexa y se presentan alguna de sus propiedades más importantes, que pueden utilizarse para desarrollar condiciones sobre optimización y esquemas computacionales adecuados para problemas de optimización.

Definición 1.29 (función convexa) *Diremos que la función $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, siendo 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 $\iff$

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

Definición 1.30 (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

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

Definición 1.31 *Se dice que f* estrictamente cóncava $\iff g = -f$ *es estrictamente convexa.*

Observación 1.32 ***Propiedades de las funciones convexas:***

1. *La cuerda que une dos puntos sobre la gráfica de una función convexa siempre está por encima de la gráfica.*
2. *La derivada de una función convexa es creciente, o al menos no decreciente.*
3. *La segunda derivada de una función convexa es siempre no negativa para todos los puntos del intervalo.*
4. *La aproximación lineal de una función convexa en cualquier punto del intervalo, siempre subestima el verdadero valor de la función.*

Proposición 1.33 *Si $f_1(\mathbf{x})$ y $f_2(\mathbf{x})$ son dos funciones convexas definidas sobre un subconjunto convexo no vacío $\Omega \subseteq \mathbb{R}^n \Longrightarrow (f_1 + f_2)(\mathbf{x})$ es convexa sobre Ω.*

Proposición 1.34 *Si $f(\mathbf{x})$ es convexa sobre $\Omega \subseteq \mathbb{R}^n$, con Ω un conjunto convexo no vacío $\Longrightarrow \forall \alpha \geq 0$, $(\alpha f)(\mathbf{x})$ es convexa sobre Ω.*

Proposición 1.35 *Si f es convexa en $\Omega \subseteq \mathbb{R}^n$ convexo no vacío $\Longrightarrow \Gamma_c = \{\mathbf{x} \in \Omega \mid f(\mathbf{x}) \leq c\}$ es convexo $\forall c \in \mathbb{R}$.*

Proposición 1.36 *Si la familia de funciones $\{f_k\}_{k=1,\ldots,m}$ son convexas sobre Ω convexo no vacío $\Longrightarrow$ Si definimos Γ_c^k como*

$$\Gamma_c^k = \{\mathbf{x} \in \Omega \mid f_k(\mathbf{x}) \leq c_k\}$$

entonces el conjunto

$$\Gamma = \cap_{k=1}^m \Gamma_c^k = \{\mathbf{x} \in \Omega \mid \mathbf{x} \in \Gamma_c^k, \forall k\}$$

es convexo.

La demostración de cada una de las proposiciones anteriores se hace directamente empleando la definición de función convexa y se deja como ejercicio.

Sea $\Omega \subseteq \mathbb{R}^n$ convexo no vacío y $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, se define el epígrafe de la función como el conjunto

$$\operatorname{epi}(f) = \{(\mathbf{x}, y) \in \mathbb{R}^n \times \mathbb{R} \mid f(\mathbf{x}) \leq y,\ \mathbf{x} \in \Omega,\ y \in \mathbb{R}\}$$

Tendremos se tiene el siguiente resultado:

Proposición 1.37 *Sea $\Omega \subseteq \mathbb{R}^n$ convexo no vacío y $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, entonces*

$$f(\mathbf{x}) \text{ convexa } \Leftrightarrow \operatorname{epi}(f) \text{ es un conjunto convexo}$$

Demostración: Supongamos en primer lugar que $f(\mathbf{x})$ es convexa en Ω. Sean $(\mathbf{x}_1, y_1)$, $(\mathbf{x}_2, y_2) \in \operatorname{epi}(f)$, y tomamos ahora para $\lambda \in [0,1]$ el punto

$$\lambda(\mathbf{x}_1, y_1) + (1-\lambda)(\mathbf{x}_2, y_2) = (\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda y_1 + (1-\lambda) y_2)$$

cumple

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2) \leq \lambda y_1 + (1-\lambda) y_2$$

luego

$$(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda y_1 + (1-\lambda) y_2) \in \operatorname{epi}(f)$$

Para demostrar el recíproco hay que tener en cuenta que si

$$\mathbf{x}_1, \mathbf{x}_2 \in \Omega \Rightarrow (\mathbf{x}_1, f(\mathbf{x}_1)), (\mathbf{x}_2, f(\mathbf{x}_2)) \in \operatorname{epi}(f)$$

de donde por la convexidad de $\operatorname{epi}(f)$ se obtiene

$$\lambda(\mathbf{x}_1, f(\mathbf{x}_1)) + (1-\lambda)(\mathbf{x}_2, f(\mathbf{x}_2)) = (\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)) \in \operatorname{epi}(f)$$

y utilizando su definición

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda) f(\mathbf{x}_2)$$

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

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 nuevas caracterización más sencillas de aplicar. Los siguientes resultados proporcionan caracterizaciones para funciones convexas diferenciables en términos del gradiente y del Hessiano.

Proposición 1.38 (Caracterización de primer orden para funciones convexas) *Sea $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, con Ω un conjunto convexo y $f(\mathbf{x}) \in \mathcal{C}^1(\Omega)$, 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 doble implicación.

$\Longrightarrow$ Si $f(\mathbf{x})$ es convexa $\Longrightarrow \forall \lambda \in [0,1]$ se tiene

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

Si asumimos $\lambda \neq 0$

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

y si ahora $\lambda \longrightarrow 0$ y tenemos en cuenta que $f \in \mathcal{C}^1(\Omega) \Longrightarrow \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$

$\Longleftarrow$ 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 ahora 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

$$\mathbf{y} = \mathbf{x}^1 \quad \Longrightarrow \quad f(\mathbf{x}^1) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{x}^1 - \mathbf{x})$$

$$\mathbf{y} = \mathbf{x}^2 \quad \Longrightarrow \quad f(\mathbf{x}^2) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{x}^2 - \mathbf{x})$$

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

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

lo que demuestra que $f(\mathbf{x})$ es convexa $\blacksquare$.

Proposición 1.39 (Caracterización de segundo orden para funciones convexas) *Sea* $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, *con* Ω *un conjunto abierto convexo no vacío y* $f(\mathbf{x}) \in \mathcal{C}^2(\Omega)$, *entonces:*

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

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 $\forall\ \mathbf{x} \in \nleq$ ocurre $\mathbf{x}^T \mathbf{H}f(\mathbf{x}^*)\mathbf{x} \geq 0$.

Puesto que $\Omega \neq \emptyset$ y abierto, entonces para $\mathbf{x} \in \mathbb{R}^n$,ocurre $\mathbf{x}^* + \lambda\mathbf{x} \in \Omega$, tomando $|\lambda| \neq 0$ y suficientemente pequeño. Si utilizamos la proposición 1.38 de caracterización de primer orden de funciones convexas por una parte y el teorema de Taylor por otra, se obtienen las siguientes expresiones:

$$f(\mathbf{x}^* + \lambda\mathbf{x}) \geq f(\mathbf{x}^*) + \lambda \nabla f(\mathbf{x}^*)\mathbf{x}$$

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

donde $o(\lambda) = \lambda^2 ||\mathbf{x}||^2 \alpha(\mathbf{x}^*, \lambda\mathbf{x})$.

Si se restan ambas expresiones obtenemos

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

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

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

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

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

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

que resulta ser la caracterización de primer orden para funciones convexas de la proposición 1.38 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 1.40 *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.

Supóngamos 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

$$\lambda\mathbf{x}^* + (1-\lambda)\,\mathbf{x}^* = \mathbf{x}^* \in \Gamma$$

Supondremos, por tanto, 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 la convexidad de $f(\mathbf{x})$ ocurre

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

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 $f(\mathbf{x})$ y supongamos que no es un mínimo 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 $\lambda, (1-\lambda) \geq 0$

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

Para un valor de λ suficientemente pequeño se contradice el hecho de que $\mathbf{x}^*$ sea un mínimo relativo y llegamos a una contradicción que aparece al suponer que $\mathbf{x}^*$ no es un mínimo global ■.

Teorema 1.41 *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 1.38) 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 por la proposición anterior es un mínimo global ■.

Proposición 1.42 *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 1.43 *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 1.44 *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: Como f es convexa

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

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

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

■

Lema 1.45 *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 1.40 es un mínimo global. Si suponemos ahora que existe otro punto $\mathbf{z}^*$ mínimo de $f(\mathbf{x})$, por el lema 1.44 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 ■.