

Capítulo 1

Fundamentos de Optimización Estática

1.1 Introducción

La teoría de optimización o programación matemática está constituida por un conjunto de resultados y métodos 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. A partir del valor de una función, llamada *función objetivo*, que está diseñada para cuantificar el rendimiento y medir la calidad de la decisión, se obtienen valores para un cierto número de variables, de forma que minimicen o maximicen esa función objetivo. Por lo general, estas variables están relacionadas entre sí mediante expresiones matemáticas o *restricciones* que limitan la elección de esos valores.

En general 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 algebraicas (las restricciones del problema), proporcionan el mejor (mayor o menor) valor posible para una función que es utilizada para medir el rendimiento del sistema en estudio. Buscamos, en resumen, valores que cumplan unas condiciones y minimicen o maximicen una función que caracteriza el sistema.

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

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

El empleo del término “*Optimizar*” en el esquema anterior, incluye a ambos objetivos de *Minimización* y/o *Maximización* de la función objetivo $f(\mathbf{x})$, cuando los puntos \mathbf{x} satisfacen el conjunto de *Restricciones*. En cualquier caso, un problema con objetivo de maximización siempre se puede transformar en otro equivalente con objetivo de minimización, y viceversa, multiplicando para ello la función objetivo por (-1) .

1.2 Elementos de Optimización

A continuación se describen los elementos necesarios para la formulación correcta de un problema de optimización de forma más detallada.

1.2.1 Función objetivo

El primer 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. En muchas aplicaciones de ingeniería se suele elegir un criterio económico, no obstante, 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, máxima carga y 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 de la función objetivo elegida.

Es importante hacer notar que sólo se utilizará un criterio de optimización para definir el óptimo. No será posible 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. 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 y son tratados como restricciones del problema.

1.2.2 Variables independientes

El segundo 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 significativo sobre la función objetivo.

Las variables independientes también se denominan variables de decisión.

1.2.3 Restricciones

Una vez determinadas la función objetivo y 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.

Ejemplo 1.1 Distancia más corta entre dos curvas: El problema consiste en 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. 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

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

o más concretamente como

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

siendo

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

las coordenadas de los puntos.

Ejemplo 1.2 Construcción de la caja de volumen máximo. El problema consiste en determinar las dimensiones de una caja rectangular de forma que contenga un volumen máximo utilizando una cantidad de material fija. El problema se puede plantear como

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

Si las dimensiones de la caja rectangular son x , y y z , entonces el problema puede plantearse como

$$\begin{array}{ll} \text{Maximizar} & xyz \\ \text{s.a.} & 2x + 2y + 2z = A \\ & x, y, z \geq 0 \end{array}$$

Ejemplo 1.3 Problema de la bombilla. El problema consiste en determinar la altura a la que se debe colocar una bombilla para que ilumine con la mayor intensidad posible una zona circular de radio dado. El problema se plantea suponiendo que la intensidad luminosa es función de la altura, $I(h)$, para establecer que

$$\begin{array}{ll} \text{Maximizar} & I(h) \\ \text{s.a.} & 0 \leq h \leq h_{\max} \end{array}$$

Ejemplo 1.4 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:

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

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

1.3 Tipos de Problemas

Podemos dividir los problemas de optimización estática según diversos criterios entre los que podemos destacar: tipo de funciones empleadas en el problema o tipo de variables utilizadas en el problemas.

1.3.1 Clasificación de problemas según el tipo de funciones

En este sentido, se distinguen dos tipos de problemas: *problemas de programación lineal* y *problemas de programación no lineal*.

Problemas de programación lineal

En este tipo de problema tratamos problemas basados en optimizar funciones objetivo que son lineales en las incógnitas y que están sujetos a restricciones que también son funciones lineales de las incógnitas. Es el tipo de problema con restricciones más sencillo de resolver. Los problemas de programación lineal son los más fáciles de plantear y también por supuesto de resolver y aunque la mayoría de problemas de interés son no lineales, a veces, como aproximación inicial, es posible definirlos como lineales, en otras ocasiones se pueden utilizar este tipo de problemas como un paso previo a la solución de problemas no lineales más complejos. El problema lineal de tipo general tiene la siguiente estructura

$$\begin{array}{ll} \text{Optimizar} & c_1x_1 + \cdots + c_nx_n \\ \text{s.a.} & \\ & a_{1,1}x_1 + \cdots + a_{1,n}x_n = b_1 \\ & \vdots \\ & a_{p,1}x_1 + \cdots + a_{p,n}x_n = b_p \\ & a_{p+1,1}x_1 + \cdots + a_{1,n}x_n \leq b_{p+1} \\ & \vdots \\ & a_{q,1}x_1 + \cdots + a_{q,n}x_n \leq b_q \\ & a_{q+1,1}x_1 + \cdots + a_{q+1,n}x_n \leq b_{q+1} \\ & \vdots \\ & a_{m,1}x_1 + \cdots + a_{m,n}x_n \leq b_m \end{array}$$

Problemas de programación no lineal

En este tipo de problemas o bien la función objetivo del problema, o bien alguna de las restricciones que limitan al problema es una función no lineal en sus argumentos. Para esta clase de problemas se pueden distinguir dos subtipos: *problemas sin restricciones* y *problemas con restricciones*. Se resume a continuación

las características básicas de cada uno. Por las características de las funciones lineales esta distinción no tiene sentido realizarla en problemas los problemas lineales descritos en la sección anterior.

1. **Problemas sin restricciones:** Aunque plantear un problema sin restricciones parece en principio sin sentido y carente de aplicación práctica, sucede sin embargo lo contrario. En primer lugar porque si ampliamos el ámbito de un problema con restricciones, aquellas que representan delimitaciones artificiales desaparecen; por ejemplo, si una de las restricciones se refiere al presupuesto de un proyecto, podemos aumentar dicho presupuesto utilizando un préstamo bancario, incorporando posteriormente a la función objetivo el reflejo del coste de ese préstamo. La segunda razón es que para su resolución, podemos convertir problemas con restricciones en problemas sin ellas. Por ejemplo el efecto de las restricciones de igualdad es el de limitar los grados de libertad del sistema, sobre todo al poner unas variables en función de otras. Podemos a veces determinar esas dependencias explícitamente de tal forma que podemos construir otro problema cuyo número de variables sea igual al verdadero grado de libertad. Por ejemplo una restricción de la forma $x_1 + x_2 = B$ puede eliminarse si sustituimos x_2 por $B - x_1$, en todas las ocasiones en las que aparece.

El estudio de problemas sin restricciones proporciona una componente clave para el caso general de problemas con restricciones. Muchos aspectos de la teoría y de los algoritmos se motivan y verifican de forma más natural para el caso sin restricciones, antes de entrar en el problema con restricciones.

2. **Problemas con restricciones:** La mayoría de los problemas prácticos se formulan como problemas con restricciones. Esto se debe a que en la mayoría de los casos un problema complejo, como es el diseño de un dispositivo, no puede tratarse libremente teniendo en cuenta todas las elecciones posibles, sino que debe descomponerse en subproblemas separados. Cada uno de estos subproblema tiene restricciones que han sido impuestas para limitar su campo de aplicación. El enunciado general de un problema de programación matemática con restricciones podría ser:

$$\begin{array}{ll} \text{Optimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & h_i(\mathbf{x}) = 0 \quad i = 1, \dots, m \\ & g_j(\mathbf{x}) \leq 0 \quad j = 1, \dots, n \\ & \mathbf{x} \in A \end{array}$$

Donde $\mathbf{x} = (x_1, \dots, x_n)$ son las variables de decisión, $f(\mathbf{x})$ es la función objetivo, $A \subseteq \mathbb{R}^n$, $h_i(\mathbf{x})$ son funciones que representan las restricciones de igualdad mientras que $g_j(\mathbf{x})$ representa el conjunto de la restricciones de desigualdad. El hecho de que solamente aparezcan restricciones del tipo $g_j(\mathbf{x}) \leq 0$, y no aparezcan restricciones del tipo $g(\mathbf{x}) \geq 0$, es debido a que estas últimas pueden transformarse en las primeras multiplicando la desigualdad por -1 .

En principio las funciones implicadas en el problema no tienen porqué tener ninguna propiedad particular, pero en nuestro caso vamos a introducir hipótesis adicionales que nos ayuden 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; generalmente $A = \mathbb{R}^n$.

1.3.2 Clasificación de problemas según el tipo de variable

Optimización continua

En los problemas continuos las variables de decisión pueden tomar cualquier valor intermedio. Dentro de estos podemos distinguir entre **problemas sin restricciones** y **problemas con restricciones**. Dentro de los problemas con restricciones diferenciamos entre problemas lineales y problemas no lineales.

Optimización discreta

En los problemas de optimización discreta las variables solamente pueden tomar valores concretos, como los problemas de **programación entera** para los cuales las variables solamente pueden tomar valores dentro del conjunto \mathbb{Z} de los números enteros.

1.4 Definiciones

El planteamiento general de un problema de optimización es el siguiente:

$$\begin{array}{ll} \text{Optimizar} & f(\mathbf{x}) \\ \text{Sujeto a} & \mathcal{R} \\ & f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R} \end{array} \quad (1.1)$$

Resolver este problema consiste, en primer lugar, en encontrar valores para la variable vectorial \mathbf{x} que resuelvan el sistema formado por las restricciones \mathcal{R} , y en segundo lugar encontrar de entre estos valores aquel o aquellos que optimicen la función real de variables reales $f(\mathbf{x})$, definida sobre un conjunto $A \subseteq \mathbb{R}^n$ en \mathbb{R} .

El conjunto \mathcal{R} representa las limitaciones del sistema, y que generalmente se expresan a través de ecuaciones e inecuaciones en la variable \mathbf{x} .

Se dan a continuación un conjunto de definiciones necesarias cuando se trata con problemas de optimización matemática.

Definición 1.5 (solución factible) Diremos que $\bar{\mathbf{x}} \in A$, es una solución factible del problema 1.1 si cumple el conjunto de las restricciones \mathcal{R} i.e. si

Definición 1.6 (conjunto factible) Dado el problema general de optimización 1.1, se define región o conjunto factible al conjunto de todas las soluciones factibles

$$\Omega = \{\bar{\mathbf{x}} \in A \mid \bar{\mathbf{x}} \text{ es una solución factible}\}$$

Definición 1.7 (solución óptima) Diremos que $\mathbf{x}^* \in A$, es una solución óptima del problema 1.1 si \mathbf{x}^* es factible y además

$$\begin{aligned} f(\mathbf{x}^*) &\leq f(\bar{\mathbf{x}}) \quad (\text{Minimización}) \\ &\quad \text{ó} \\ f(\mathbf{x}^*) &\geq f(\bar{\mathbf{x}}) \quad (\text{Maximización}) \end{aligned}$$

para cualquier otra solución factible $\forall \bar{\mathbf{x}} \in \Omega$, con $\bar{\mathbf{x}} \neq \mathbf{x}^*$.

Nuestro objetivo al intentar resolver un problema de optimización del tipo descrito en 1.1, será encontrar la solución óptima del mismo, es decir, la “mejor” de las soluciones factibles. Los valores de las soluciones óptimas son generalmente conocidos como *extremos globales* del problema y se habla de *mínimo* o *máximo global* dependiendo del objetivo del problema de optimización y que pueden definirse como

Definición 1.8 (mínimo global) Un punto $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ es un punto de mínimo global o mínimo global débil del problema 1.1 si:

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

Será mínimo global estricto ó mínimo global fuerte si la desigualdad es estricta: $f(\mathbf{x}^*) < f(\bar{\mathbf{x}})$.

Definición 1.9 (máximo global) Un punto $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ es un punto de máximo global ó máximo global débil del problema 1.1

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

Será un máximo global estricto ó máximo global fuerte si la desigualdad es estricta.

Resolver un problema de optimización es encontrar, si existen, el extremo o extremos globales de $f(\mathbf{x})$ sobre el conjunto Ω . Desde el punto de vista práctico y computacional en algunas ocasiones bastará con obtener los llamados extremos locales y que definiremos a continuación:

Definición 1.10 (mínimo relativo) Diremos que un punto $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ es un punto de mínimo relativo o local del problema 1.1 si:

$$\exists \epsilon > 0 \quad \forall \bar{\mathbf{x}} \in \Omega; \quad \bar{\mathbf{x}} \neq \mathbf{x}^*; \quad \|\bar{\mathbf{x}} - \mathbf{x}^*\| < \epsilon \implies f(\mathbf{x}^*) \leq f(\bar{\mathbf{x}})$$

donde Ω es el conjunto factible del problema.

A este tipo de mínimos también se les suele llamar *mínimos locales débiles*, para diferenciarlos de los *mínimos locales fuertes*, para los cuales se cumple la desigualdad estricta del tipo $f(\mathbf{x}^*) < f(\bar{\mathbf{x}})$, y que también reciben el nombre de *mínimo relativo o local estricto*.

Definición 1.11 (máximo relativo) Diremos que un punto $\mathbf{x}^* \in \Omega \subseteq \mathbb{R}^n$ es un punto de máximo relativo o local del problema 1.1 si:

$$\exists \epsilon > 0 \text{ } \bigwedge \bar{\mathbf{x}} \in \Omega; \quad \bar{\mathbf{x}} \neq \mathbf{x}^*; \quad \|\bar{\mathbf{x}} - \mathbf{x}^*\| < \epsilon \implies f(\mathbf{x}^*) \geq f(\bar{\mathbf{x}})$$

Si la desigualdad es estricta el punto se llama máximo relativo estricto. Análogamente al caso de mínimo llamaremos al primero máximo local débil, mientras que el segundo será un máximo local fuerte.

Un punto \mathbf{x}^* que sea máximo o mínimo local de una función, se dice que es un *extremo local (débil o fuerte)*, mientras que si es un máximo o mínimo global entonces será un *extremo global (débil o fuerte)*.

La teoría inicial asociada a la optimización, especialmente en programación no lineal, propone más que la computación, 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, 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. **La cuestión estática:** ¿Cómo podemos determinar si un punto \mathbf{x}^* es o no la solución óptima?
2. **La 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}^* ?

En este texto y para cada tipo de optimización (lineal y no lineal, con y sin restricciones) se comienza por la primera cuestión (cuestión estática) desarrollando un conjunto de *criterios de optimización* con el fin de determinar si una solución dada es óptima. A continuación se trata la segunda cuestión, construyendo algoritmos y métodos numéricos para encontrar la solución a partir de un punto inicial dado.

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 1.12 (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

$$\begin{aligned} f(\mathbf{x}^{\min}) &= \min_{\mathbf{x} \in K} f(\mathbf{x}) \\ \exists \mathbf{x}^{\min}, \mathbf{x}^{\max} \in K &\Rightarrow \\ f(\mathbf{x}^{\max}) &= \max_{\mathbf{x} \in K} f(\mathbf{x}) \end{aligned}$$

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.

1.5 Convexidad

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 se puede garantizar que a partir de un óptimo local de un problema podamos alcanzar el óptimo global del mismo, que realmente es lo que nos interesa.

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 .

1.5.1 Conjuntos convexos y envoltura convexa

Definición 1.13 (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 1.14 (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.

Por convenio el conjunto vacío \emptyset es convexo.

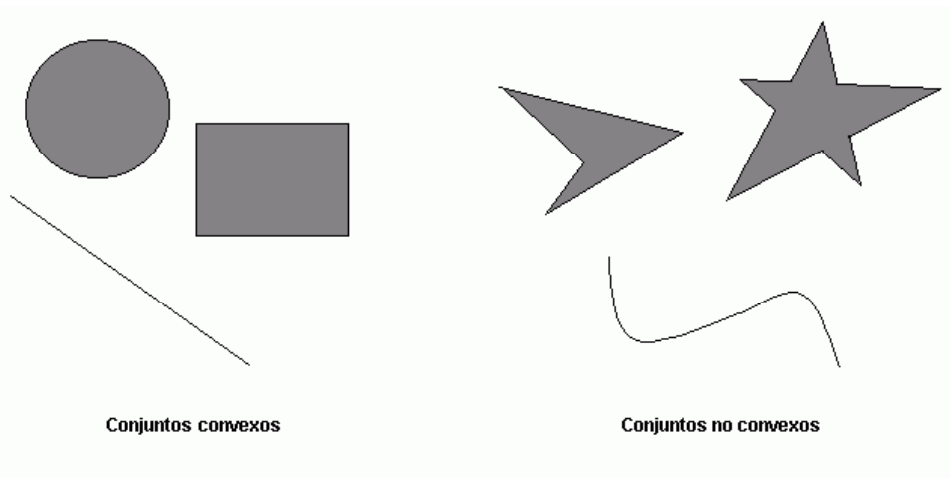


Figura 1.1: Convexidad en \mathbb{R}^2 .

La figura 1.1 representa algunos conjuntos convexos y no convexos de \mathbb{R}^2 .

Uno de los tipos más importantes de conjunto convexo es el *hiperplano*. Los hiperplanos dominan la teoría de optimización, presentándose en forma de multiplicadores de Lagrange, teoría de dualidad o cálculos del gradiente.

Definición 1.15 (hiperplano) *Si $\mathbf{a} \in \mathbb{R}^n$; $\mathbf{a} \neq \mathbf{0}$, $c \in \mathbb{R}$. 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 1.16 (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:*

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

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

y semiespacios abiertos positivos y negativos a:

$$\overset{\circ}{H}_+ = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T \mathbf{x} > b\}$$

$$\overset{\circ}{H}_- = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T \mathbf{x} < b\}$$

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 algebraica de dos conjuntos convexos también es convexa. Su demostración se deja como ejercicio.

Lema 1.17 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

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

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 intersección de m semiespacios.

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

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

también es un conjunto convexo.

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

Proposición 1.18 El conjunto imagen de un conjunto convexo $\Omega \subseteq \mathbb{R}^n$, mediante una aplicación lineal $f : \Omega \rightarrow \mathbb{R}^m$, no nula definida en él es un conjunto convexo.

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

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

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

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

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

$$\exists \mathbf{x}^1 \in \Omega : \mathbf{y}^1 = f(\mathbf{x}^1)$$

$$\exists \mathbf{x}^2 \in \Omega : \mathbf{y}^2 = f(\mathbf{x}^2)$$

Se construye el segmento que une \mathbf{x}^1 y \mathbf{x}^2 y se evalúa la función en él, teniendo en cuenta su propiedad de linealidad:

$$\lambda \mathbf{x}^1 + (1 - \lambda) \mathbf{x}^2 \Rightarrow f(\lambda \mathbf{x}^1 + (1 - \lambda) \mathbf{x}^2) = \lambda f(\mathbf{x}^1) + (1 - \lambda) f(\mathbf{x}^2) = \lambda \mathbf{y}^1 + (1 - \lambda) \mathbf{y}^2 \in f(\Omega)$$

y por tanto el segmento que une las imágenes está en $f(\Omega)$.

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

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

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

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

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 1.20 (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

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

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

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

donde m es un entero positivo y $\{\mathbf{x}^1, \dots, \mathbf{x}^m\} \in \Omega$. 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 $\mathbf{x} \in H(\Omega)$. 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 1.2 un conjunto no convexo y su envoltura convexa.

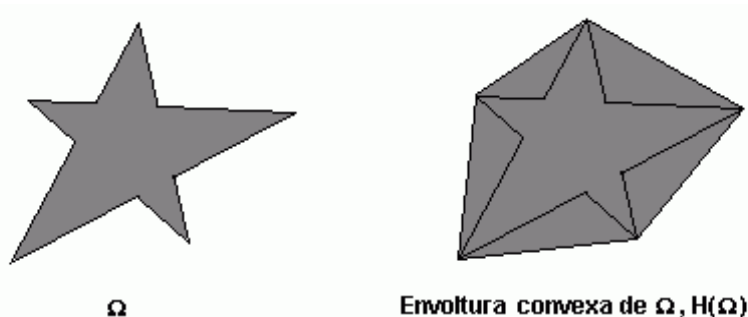


Figura 1.2: Envoltura convexa de un conjunto.

Definición 1.21 (polígono, simplex, poliedro) La envoltura convexa de un conjunto finito de puntos de \mathbb{R}^n , $\Omega = \{\mathbf{x}^1, \dots, \mathbf{x}^{k+1}\}$, se denomina polígono. Si además el conjunto $\{\mathbf{x}^2 - \mathbf{x}^1, \mathbf{x}^3 - \mathbf{x}^1, \dots, \mathbf{x}^{k+1} - \mathbf{x}^1\}$ es linealmente independiente, entonces esta envoltura convexa $H(\Omega) = H(\mathbf{x}^1, \dots, \mathbf{x}^{k+1})$ se llama simplex de vértices $\mathbf{x}^1, \dots, \mathbf{x}^{k+1}$. Por último un polígono acotado se denomina poliedro.

Es posible demostrar que un polígono es la intersección de un número finito de semiespacios cerrados, equivalentemente a un conjunto de desigualdades lineales de la forma:

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

Definición 1.22 (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 $\Omega \iff$

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

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

$$\Omega = \{ \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \mid 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 2 \}$$

tiene 4 puntos extremos dados por $(0, 0)$, $(0, 2)$, $(2, 0)$ y $(2, 2)$ (Figura 1.3)

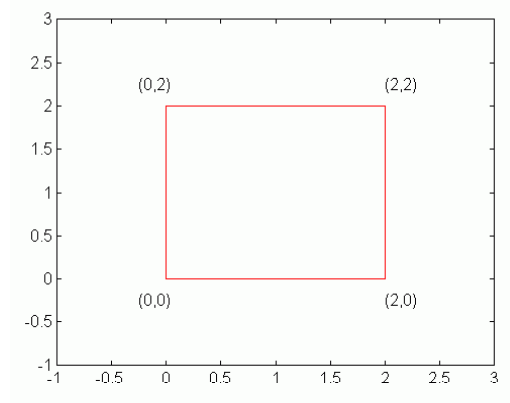


Figura 1.3: Puntos extremos.

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 .

1.6 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.

1.6.1 Definiciones

Definición 1.23 (función convexa) Diremos que la función $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, 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

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

Definición 1.24 (función cóncava) Diremos que la función $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, 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 1.25 Se dice que f estrictamente cóncava $\iff g = -f$ es estrictamente convexa.

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 1.26 Si $f_1(\mathbf{x})$ y $f_2(\mathbf{x})$ son dos funciones convexas definidas sobre el subconjunto convexo no vacío $\Omega \subseteq \mathbb{R}^n \implies (f_1 + f_2)(\mathbf{x})$ es convexa.

Proposición 1.27 Si $f(\mathbf{x})$ es convexa sobre $\Omega \subseteq \mathbb{R}^n$, un conjunto convexo no vacío $\implies \forall \alpha \geq 0, (\alpha f)(\mathbf{x})$ es convexa sobre Ω .

Proposición 1.28 Si f es convexa en $\Omega \subseteq \mathbb{R}^n$ convexo no vacío $\implies \Gamma_c = \{\mathbf{x} \in \Omega \mid f(\mathbf{x}) \leq c\}$ es convexo $\forall c \in \mathbb{R}$.

Proposición 1.29 Si la familia de funciones $\{f_k\}_{k=1,\dots,m}$ son convexas sobre Ω convexo no vacío \implies Si definimos Γ_c^k como

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

entonces el conjunto

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

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.

Proposición 1.30 Sea $\Omega \subseteq \mathbb{R}^n$ convexo no vacío y $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. Entonces se define el epígrafe de la función como el conjunto

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

entonces se tiene el siguiente resultado:

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

Demostración: Supongamos en primer lugar que $f(\mathbf{x})$ es convexa en Ω . Sean $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2) \in \text{epi}(f)$, 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

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

Para demostrar el recíproco hay que tener en cuenta que si

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

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

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

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 con el fin de comprobar si una función es o no convexa y 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 1.31 (Teorema de caracterización de primer orden para funciones convexas) Sea $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con Ω un conjunto convexo y $f(\mathbf{x}) \in \mathcal{C}^1(\Omega)$, entonces:

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

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

Demostración: Demostramos la doble implicación.

“ \implies ”

Si $f(\mathbf{x})$ es convexa $\implies \forall \lambda \in [0, 1]$ 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 $\lambda \rightarrow 0$ y tenemos en cuenta que $f \in \mathcal{C}^1(\Omega) \implies \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$

“ \impliedby ”

Supongamos ahora que $\nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \Omega$

Si $\mathbf{x}^1, \mathbf{x}^2 \in \Omega$, $\lambda \in [0, 1] \implies \mathbf{x} = \lambda \mathbf{x}^1 + (1 - \lambda)\mathbf{x}^2 \in \Omega$, por ser Ω un conjunto convexo. Si utilizamos la desigualdad anterior 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

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

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

Si ahora multiplicamos por λ la primera desigualdad y por $(1 - \lambda)$ la segunda, y sumamos el resultado

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

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

Proposición 1.32 (Teorema de 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:

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

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: Supongamos que $f(\mathbf{x})$ es convexa en Ω y sea $\mathbf{x}^* \in \Omega$. Hay que demostrar que $\mathbf{x}^T \mathbf{H}f(\mathbf{x}^*) \mathbf{x} \geq 0; \forall \mathbf{x} \in \mathbb{R}^n$.

Puesto que $\Omega \neq \emptyset$ y abierto, entonces para $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^* + \lambda \mathbf{x} \in \Omega$, con $|\lambda| \neq 0$ y suficientemente pequeño. Si utilizamos la proposición 1.31 de caracterización de funciones convexas por una parte y el teorema de Taylor por otra, se obtienen las siguientes expresiones:

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

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

Donde $o(\lambda) = \lambda^2 \|\mathbf{x}\|^2 \alpha(\mathbf{x}^*, \lambda \mathbf{x})$. Si se restan ambas expresiones obtenemos:

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

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

Recíprocamente si suponemos ahora que la matriz Hessiana de f 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 $\mathbf{x}^{**} = \lambda \mathbf{x}^* + (1 - \lambda) \mathbf{x} \in \Omega$, $\lambda \in (0, 1)$ y $\mathbf{H}f(\mathbf{x}^{**})$ semidefinida positiva. Lo que nos conduce a que $(\mathbf{x} - \mathbf{x}^*)^T \mathbf{H}f(\mathbf{x}^{**}) (\mathbf{x} - \mathbf{x}^*) \geq 0$ y se concluye

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

que es la caracterización de funciones convexas dada en la proposición 1.31 y f es convexa en Ω ■.

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.

1.7 Optimización de funciones convexas

En esta sección consideraremos el problema de minimizar y maximizar una función convexa sobre un conjunto convexo y se desarrollarán las condiciones necesarias y/o suficientes para la optimización bajo hipótesis de convexidad.

Teorema 1.33 Sea $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ con Ω un conjunto convexo. Si $f(\mathbf{x})$ es convexa \implies El conjunto Γ donde f alcanza su mínimo es convexo y cualquier mínimo local de $f(\mathbf{x})$ es un mínimo global.

Demostración: Si $\Gamma = \emptyset$, es decir f no tiene mínimos relativos, entonces el teorema se cumple por la convexidad del conjunto vacío.

Supongamos ahora que c_0 es el mínimo de f . Entonces, el conjunto $\Gamma = \{\mathbf{x} \in \Omega \mid f(\mathbf{x}) \leq c_0\}$ es convexo por la proposición 1.28.

Supongamos ahora que $\mathbf{x}^* \in \Omega$ es un mínimo local de f 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 en el conjunto convexo Ω obtenemos 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 $f(\mathbf{y}) < f(\mathbf{x}^*)$ 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 producida al suponer que \mathbf{x}^* no es un mínimo global ■.

Del mismo modo podemos dar un teorema similar para funciones cóncavas:

Teorema 1.34 Sea $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ con Ω un conjunto convexo. Si $f(\mathbf{x})$ es cóncava \implies El conjunto Γ donde f alcanza su máximo es convexo y cualquier máximo local de $f(\mathbf{x})$ es un máximo global..

Teorema 1.35 Sea $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ con Ω un conjunto convexo con $f \in C^1(\Omega)$ una función convexa. Si existe $\mathbf{x}^* \in \Omega$ que cumple la siguiente propiedad

$$\forall \mathbf{y} \in \Omega \Rightarrow \nabla f(\mathbf{x}^*)^T (\mathbf{y} - \mathbf{x}^*) \geq 0$$

Entonces \mathbf{x}^* es un punto de mínimo 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 1.31) 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 1.36 Si Ω es un subconjunto no vacío, convexo y compacto de \mathbb{R}^n cuyo conjunto de puntos extremos es finito y si $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa sobre $\Omega \Rightarrow f(\mathbf{x})$ posee en Ω un máximo global y se encuentra en uno de sus puntos extremos

Proposición 1.37 Si Ω es un subconjunto no vacío, convexo y compacto de \mathbb{R}^n cuyo conjunto de puntos extremos es finito y si $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ es una función cóncava sobre $\Omega \Rightarrow f(\mathbf{x})$ posee en Ω un mínimo global y se encuentra en uno de sus puntos extremos..

Teorema 1.38 Sea $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ convexa en Ω convexo y compacto \Rightarrow Si $f(\mathbf{x})$ tiene máximo en Ω entonces lo alcanza en un punto extremo de Ω .

Demostración: Supongamos que f alcanza en \mathbf{x}^* su máximo global. Si $\mathbf{x}^* \notin \partial\Omega$, consideremos L una recta que pasa por \mathbf{x}^* . $L \cap \Omega$ es el segmento que pasa por \mathbf{x}^* , de extremos $\mathbf{y}^1, \mathbf{y}^2 \in \partial\Omega$. Como \mathbf{x}^* pertenece al segmento, existirá $\lambda \in (0, 1)$ con $\mathbf{x}^* = \lambda\mathbf{y}^1 + (1 - \lambda)\mathbf{y}^2$. Al ser $f(\mathbf{x})$ una función convexa en Ω :

$$f(\mathbf{x}^*) \leq \lambda f(\mathbf{y}^1) + (1 - \lambda)f(\mathbf{y}^2) \leq \max\{f(\mathbf{y}^1), f(\mathbf{y}^2)\}$$

Como $f(\mathbf{x}^*)$ es el máximo global de f tendría que ocurrir: $f(\mathbf{x}^*) = f(\mathbf{y}^1)$ o $f(\mathbf{x}^*) = f(\mathbf{y}^2)$ y el máximo se alcanza también en \mathbf{y}^1 ó en \mathbf{y}^2 que son puntos de la frontera ■.

El mismo teorema para funciones cóncavas sería:

Teorema 1.39 Sea $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ cóncava en Ω convexo y compacto \Rightarrow Si $f(\mathbf{x})$ tiene mínimo en Ω entonces lo alcanza en un punto extremo de Ω .

Lema 1.40 Si $\mathbf{x}^*, \mathbf{y}^*$ son dos puntos de mínimo global de una función convexa $f(\mathbf{x})$, $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con Ω convexo $\Rightarrow \forall \lambda \in [0, 1]$, los puntos definidos como $\mathbf{z}^*(\lambda) = \lambda\mathbf{x}^* + (1 - \lambda)\mathbf{y}^*$ también son mínimos 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 $\Rightarrow f(\mathbf{y}^*) = f(\mathbf{x}^*)$ 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

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

■

Para funciones cóncavas

Lema 1.41 Si $\mathbf{x}^*, \mathbf{y}^*$ son dos puntos de máximo global de una función cóncava $f(\mathbf{x})$, $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con Ω convexo $\Rightarrow \forall \lambda \in [0, 1]$, los puntos definidos como $\mathbf{z}^*(\lambda) = \lambda\mathbf{x}^* + (1 - \lambda)\mathbf{y}^*$ también son máximos globales de $f(\mathbf{x})$.

Teorema 1.42 Sea $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con Ω un conjunto convexo no vacío y $f(\mathbf{x})$ convexa. Sea \mathbf{x}^* un mínimo local de f . Entonces, si \mathbf{x}^* es un mínimo local estricto de $f(\mathbf{x})$ o si $f(\mathbf{x})$ es estrictamente convexa en Ω , entonces \mathbf{x}^* es el único mínimo global de $f(\mathbf{x})$ en Ω .

Demostración: Si \mathbf{x}^* es un mínimo local entonces por el teorema 1.33 es un mínimo global. Si suponemos ahora que existe otro punto \mathbf{z}^* mínimo de $f(\mathbf{x})$, por el lema 1.40 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:

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

y por tanto \mathbf{x}^* no sería mínimo global ■

Que para funciones cóncavas podría enunciarse como:

Teorema 1.43 Sea $f : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, con Ω un conjunto convexo no vacío y $f(\mathbf{x})$ cóncava. Sea \mathbf{x}^* un máximo local de f . Entonces, si \mathbf{x}^* es un máximo local estricto de $f(\mathbf{x})$ o si $f(\mathbf{x})$ es estrictamente cóncava sobre Ω , entonces \mathbf{x}^* es el único máximo global de $f(\mathbf{x})$ en Ω .

1.8 Bibliografía Básica del Tema

1. *Engineering Optimization. Methods and Applications*. Reklaitis, G.V.; Ravindran, A. and Ragsdell K.M. **Ed John Wiley and Sons (1983)**
2. *Decisiones de Optimización*. Manuel Mocholí Arce y Ramón Sala Garrido. **Editorial Tirant lo Blanch (1996)**
3. *Teoría de la Optimización*. Vicente Novo Sanjurjo. **Aula Abierta. Editorial U.N.E.D. (1997)**
4. *Programación Matemática*. Alejandro Balbas y José Antonio Gil. **Editorial AC. (1990)**