

CONDICIONES DE KARUSH-KUHN-TUCKER

Pedro Miranda Menéndez

Resumen

En este tema vamos a realizar un estudio teórico de los problemas de Programación No Lineal. Comenzaremos con un estudio de los conjuntos convexos, que nos permitirá enunciar resultados de separación de conjuntos convexos. El siguiente paso es enunciar y demostrar el Lema de Farkas y otros resultados similares conocidos como Teoremas de Alternativa. A continuación se planteará el problema general de Programación No Lineal, comparándolo con el problema de Programación Lineal; como veremos, en el caso general hay más situaciones y es más difícil de resolver; pasaremos entonces a ver las condiciones que son necesarias para que un punto factible sea un mínimo local; desarrollando estas condiciones llegaremos a las condiciones de Fritz-John en primer lugar y luego a las condiciones de Karush-Kuhn-Tucker, que se demuestran a partir de los Teoremas de Alternativa. Sin embargo, estos resultados no permiten resolver cualquier tipo de problema de forma satisfactoria; por ello, se estudian como punto final del tema las funciones convexas, en las que se ve que sólo hay un punto de KKT y que es el óptimo; esto constituye la base de la resolución de problemas de Programación Cuadrática a partir del algoritmo de Lemke.

1. Conjuntos convexos. Teoremas de alternativa

Vamos a dar una pequeña introducción a la Teoría de Conjuntos Convexos, que resume lo visto en Investigación Operativa; los resultados sobre Análisis de Conjuntos Convexos nos servirán para introducir y desarrollar algunos teoremas de alternativa.

1.1. Conjuntos convexos

Definition 1. Dado $C \in \mathbb{R}^n$, se dice que C es un conjunto convexo si la combinación convexa de dos puntos de C está en C , es decir,

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in C, \forall \alpha \in [0, 1], \alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_2 \in C.$$

Como consecuencias inmediatas de la definición se tiene

Lema 1. Sean C_1 y C_2 dos conjuntos convexos en \mathbb{R}^n . Entonces:

- $C_1 \cap C_2$ es un conjunto convexo.
- $C_1 \oplus C_2 := \{\mathbf{x}_1 + \mathbf{x}_2 : \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2\}$ es un conjunto convexo.

Demostración:

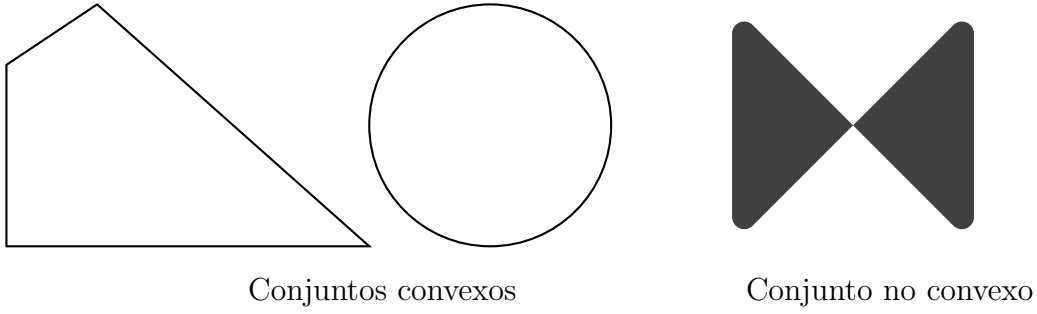


Figura 1: Ejemplos de conjuntos convexos y no convexos.

- Sean $\mathbf{x}_1, \mathbf{x}_2 \in C_1 \cap C_2$. Entonces, para $\lambda \in [0, 1]$ tenemos

$$\left. \begin{array}{l} \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2 \in C_1 \\ \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2 \in C_2 \end{array} \right\} \Rightarrow \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2 \in C_1 \cap C_2.$$

- Sean $\mathbf{x}, \mathbf{y} \in C_1 \oplus C_2$. Por definición,

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2, \mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2, \mathbf{x}_1, \mathbf{y}_1 \in C_1, \mathbf{x}_2, \mathbf{y}_2 \in C_2.$$

Entonces, para $\lambda \in [0, 1]$ tenemos

$$\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} = \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{y}_1 + \lambda \mathbf{x}_2 + (1 - \lambda) \mathbf{y}_2.$$

Como C_1 y C_2 son convexos, tenemos que $\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{y}_1 \in C_1$ y $\lambda \mathbf{x}_2 + (1 - \lambda) \mathbf{y}_2 \in C_2$, con lo que se tiene el resultado. ■

Lema 2. *Un conjunto $C \subseteq \mathbb{R}^n$ es convexo si y sólo si para cualesquiera $\mathbf{x}_1, \dots, \mathbf{x}_k \in C$ se cumple*

$$\lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k \in C, \forall \lambda_i > 0, i = 1, \dots, k, \sum_{i=1}^k \lambda_i = 1.$$

Demostración: \Rightarrow) La demostración se hace por inducción en k . Para $k = 2$ el resultado es cierto por hipótesis.

Sea $k > 2$ y supongamos el resultado cierto hasta $k - 1$. Entonces,

$$\lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k = \lambda_1 \mathbf{x}_1 + (1 - \lambda_1) \left[\frac{\lambda_2}{1 - \lambda_1} \mathbf{x}_2 + \dots + \frac{\lambda_k}{1 - \lambda_1} \mathbf{x}_k \right].$$

Aplicando el caso $k = 2$ e inducción se obtiene

$$\left. \frac{\lambda_2}{1-\lambda_1} \mathbf{x}_2 + \dots + \frac{\lambda_k}{1-\lambda_1} \mathbf{x}_k \in C \right\} \Rightarrow \lambda_1 \mathbf{x}_1 + (1-\lambda_1) \left[\frac{\lambda_2}{1-\lambda_1} \mathbf{x}_2 + \dots + \frac{\lambda_k}{1-\lambda_1} \mathbf{x}_k \right] \in C.$$

\Leftrightarrow) Trivial. ■

Ejemplo 1. Fijados $a_1, \dots, a_n \in \mathbb{R}$ no todos nulos y $b \in \mathbb{R}$, los conjuntos

$$C := \{(x_1, \dots, x_n) : a_1 x_1 + \dots + a_n x_n = b\}, \quad D := \{(x_1, \dots, x_n) : a_1 x_1 + \dots + a_n x_n \leq b\}$$

son conjuntos convexos. Por el lema anterior, esto hace que la región factible de un problema de Programación Lineal sea un conjunto convexo. El conjunto

$$F := (\cap_{i=1}^m C_i) \cap (\cap_{j=1}^l D_j),$$

donde

$$C_i := \{(x_1, \dots, x_n) : a_{i1} x_1 + \dots + a_{in} x_n = b_i\}, \quad D_j := \{(x_1, \dots, x_n) : a_{j1} x_1 + \dots + a_{jn} x_n \leq b_j\}$$

con $a_{i1}, \dots, a_{in}, a_{j1}, \dots, a_{jn}, b_i, b_j \in \mathbb{R}, \forall i, j$, es un conjunto convexo llamado **poliedro convexo**. Si además es un conjunto acotado se llama **politopo**.

Definition 2. Sea H un conjunto de \mathbb{R}^n . Se define su **envolvente convexa** como el conjunto de todas las combinaciones convexas (finitas) de puntos de H , es decir,

$$E(H) := \{\mathbf{x} : \mathbf{x} = \sum_{i=1}^r \lambda_i \mathbf{x}_i, \lambda_i > 0, \mathbf{x}_i \in H, i = 1, \dots, r, \sum_{i=1}^r \lambda_i = 1, r \in \mathbb{N}\}$$

Lema 3. Dado $H \subseteq \mathbb{R}^n$, se tiene que $E(H)$ es un conjunto convexo. Además, es el menor conjunto convexo que contiene a H . En particular, si H es convexo, entonces $E(H) = H$.

Demostración: Sean $\mathbf{x}, \mathbf{y} \in E(H)$. Entonces,

$$\begin{aligned} \mathbf{x} &= \sum_{i=1}^r \lambda_i \mathbf{x}_i, \lambda_i > 0, \mathbf{x}_i \in H, i = 1, \dots, r, \sum_{i=1}^r \lambda_i = 1, \\ \mathbf{y} &= \sum_{i=1}^s \gamma_i \mathbf{y}_i, \gamma_i > 0, \mathbf{y}_i \in H, i = 1, \dots, s, \sum_{i=1}^s \gamma_i = 1. \end{aligned}$$

Así, para $\alpha \in [0, 1]$ tenemos que

$$\alpha \mathbf{x} + (1-\alpha) \mathbf{y} = \alpha \sum_{i=1}^r \lambda_i \mathbf{x}_i + (1-\alpha) \sum_{i=1}^s \gamma_i \mathbf{y}_i = \sum_{i=1}^r \alpha \lambda_i \mathbf{x}_i + \sum_{i=1}^s (1-\alpha) \gamma_i \mathbf{y}_i.$$

Como $\sum_{i=1}^r \lambda_i + \sum_{i=1}^s (1 - \alpha) \gamma_i = 1$ y son todos no negativos se concluye que $H(S)$ es convexo.

Sea C convexo conteniendo a S . Entonces, dados $\mathbf{x}_i \in S, i = 1, \dots, r$ se tiene que por ser C convexo,

$$\sum_{i=1}^r \lambda_i \mathbf{x}_i \in C, \lambda_i > 0, \mathbf{x}_i \in S, i = 1, \dots, r, \sum_{i=1}^r \lambda_i = 1.$$

Luego $H(S) \subseteq C$. ■

1.2. Teoremas de Separación

Unas propiedades muy importantes de los conjuntos convexos tienen que ver con los llamados Teoremas de Separación. Básicamente, los Teoremas de Separación dan resultados que permiten separar mediante un hiperplano distintos conjuntos convexos.

Teorema 1. Sea $C \subseteq \mathbb{R}^n$ un conjunto convexo, no vacío y cerrado y sea $\mathbf{y} \notin C$. Entonces, existe y es único un punto $\bar{\mathbf{x}} \in C$ tal que

$$d(\mathbf{y}, \bar{\mathbf{x}}) = \min\{d(\mathbf{y}, \mathbf{z}) : \mathbf{z} \in C\}.$$

Además, dado $\mathbf{z} \in C$, se tiene que $\bar{\mathbf{x}}$ es el único punto de C que cumple

$$(\mathbf{y} - \bar{\mathbf{x}})^t (\mathbf{z} - \bar{\mathbf{x}}) \leq 0.$$

En otras palabras, el ángulo que forman los vectores $\mathbf{y} - \bar{\mathbf{x}}$ y $\mathbf{z} - \bar{\mathbf{x}}$ es mayor o igual que 90 grados.

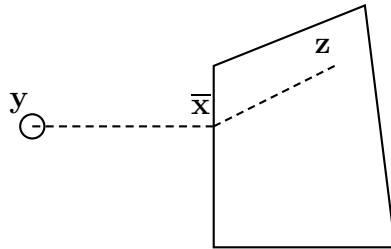


Figura 2: interpretación geométrica del Teorema 1.

Teorema 2. Sea $C \subseteq \mathbb{R}^n$ un conjunto no vacío, convexo y cerrado y sea $\mathbf{x} \notin C$. Entonces, existe un vector \mathbf{p} tal que

$$\mathbf{p}^t \mathbf{x} > \mathbf{p}^t \mathbf{y}, \mathbf{y} \in C.$$

Es decir, que podemos encontrar un vector \mathbf{p} que define la dirección de un hiperplano de forma que \mathbf{x} quede en uno de los subespacios estrictos que define el hiperplano y el conjunto C en el otro semiespacio estricto. En este caso, se habla de una **separación fuerte**.

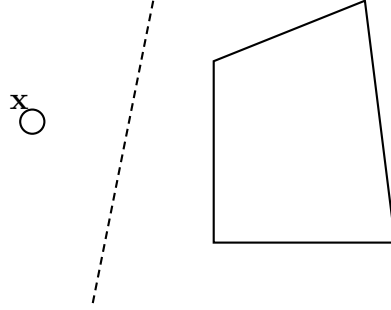


Figura 3: Interpretación geométrica del Teorema 2.

Teorema 3. Sean C_1, C_2 dos conjuntos convexos no vacíos tales que $C_1 \cap C_2 = \emptyset$. Entonces existe un vector \mathbf{p} no nulo tal que

$$\inf\{\mathbf{p}^t \mathbf{x} : \mathbf{x} \in C_1\} \geq \sup\{\mathbf{p}^t \mathbf{x} : \mathbf{x} \in C_2\}.$$

Es decir, que podemos encontrar un vector \mathbf{p} que define la dirección de un hiperplano de forma que S_1 quede en uno de los subespacios (ahora no necesariamente estrictos) que define el hiperplano y el conjunto S_2 en el otro semiespacio. En este caso, se habla de una **separación débil**.

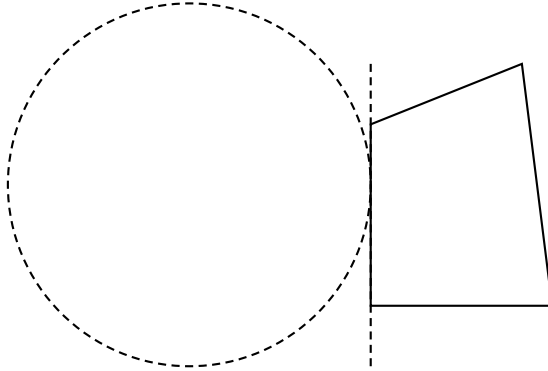


Figura 4: Interpretación geométrica del Teorema 3.

Teorema 4. Sean C_1, C_2 dos conjuntos convexos no vacíos y cerrados tales que $C_1 \cap C_2 = \emptyset$. Supongamos además que C_1 está acotado (es decir, es compacto). Entonces existe un vector \mathbf{p} no nulo y un $\epsilon > 0$ tales que

$$\inf\{\mathbf{p}^t \mathbf{x} : \mathbf{x} \in C_1\} \geq \epsilon + \sup\{\mathbf{p}^t \mathbf{x} : \mathbf{x} \in C_2\}.$$

Es decir, que podemos encontrar un vector \mathbf{p} que define la dirección de un hiperplano de forma que C_1 quede en uno de los subespacios (ahora sí estricto) que define el hiperplano y el conjunto C_2 en el otro semiespacio estricto. En este caso, se habla de una **separación estricta**.

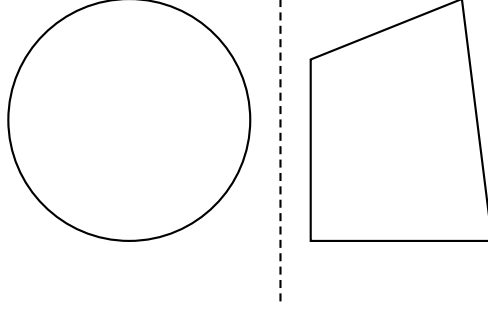


Figura 5: Interpretación geométrica del Teorema 4.

1.3. Teoremas de Alternativa

Los Teoremas de Alternativa son resultados en los que se ve que dos sistemas de inecuaciones lineales son incompatibles entre ellos. Es decir, uno y sólo uno de ellos tiene solución.

Teorema 5. (*Lema de Farkas*) Sea A una matriz $m \times n$ y sea $\mathbf{c} \in \mathbb{R}^n$. Entonces uno y sólo uno de los siguientes sistemas tiene solución:

- $\exists \mathbf{x} \in \mathbb{R}^n$ tal que $A\mathbf{x} \leq \mathbf{0}$, $\mathbf{c}^t \mathbf{x} > 0$.
- $\exists \mathbf{y} \in \mathbb{R}^m$ tal que $A^t \mathbf{y} = \mathbf{c}$, $\mathbf{y} \geq \mathbf{0}$.

Demostración: Supongamos que exista un vector \mathbf{y} tal que $\mathbf{y}^t A = \mathbf{c}^t$, $\mathbf{y} \geq \mathbf{0}$. Si existiese un vector \mathbf{x} tal que $A\mathbf{x} < \mathbf{0}$, $\mathbf{c}^t \mathbf{x} > 0$ entonces se tiene:

$$\mathbf{y}^t A\mathbf{x} = \begin{cases} \mathbf{y}^t (A\mathbf{x}) \leq 0 \\ (\mathbf{y}^t A)\mathbf{x} = \mathbf{c}^t \mathbf{x} > 0 \end{cases}$$

Y esto nos lleva a contradicción.

Supongamos que no exista \mathbf{y} tal que $\mathbf{y}^t A = \mathbf{c}^t$, $\mathbf{y} \geq \mathbf{0}$. Entonces $\mathbf{c} \notin S := \{A^t \mathbf{y} : \mathbf{y} \geq \mathbf{0}\}$. Ahora bien, S es un conjunto cerrado y convexo, y entonces podemos usar los teoremas de separación para concluir que existe un vector \mathbf{x} tal que $\mathbf{c}^t \mathbf{x} > \mathbf{y}^t A\mathbf{x}$, $\forall \mathbf{y} \geq \mathbf{0}$. En particular, si $\mathbf{y} = \mathbf{0}$, se tiene que $\mathbf{c}^t \mathbf{x} > 0$. Por otra parte, como las coordenadas del vector \mathbf{y} pueden hacerse arbitrariamente grandes, si $A\mathbf{x} \not\leq \mathbf{0}$, entonces se podría encontrar un vector \mathbf{y} tal que $\mathbf{c}^t \mathbf{x} < \mathbf{y}^t A\mathbf{x}$, lo que es una contradicción. ■

Como consecuencia de este resultados tenemos los siguientes corolarios:

Corolario 1. Sea A una matriz $m \times n$ y sea $\mathbf{c} \in \mathbb{R}^n$. Entonces uno y sólo uno de los siguientes sistemas tiene solución:

- $\exists \mathbf{x} \in \mathbb{R}^n$ tal que $A\mathbf{x} \leq \mathbf{0}$, $\mathbf{c}^t \mathbf{x} > 0$, $\mathbf{x} \geq \mathbf{0}$.
- $\exists \mathbf{y} \in \mathbb{R}^m$ tal que $A^t \mathbf{y} \geq \mathbf{c}$, $\mathbf{y} \geq \mathbf{0}$.

Demostración: El primer sistema se puede escribir como

$$\begin{pmatrix} A \\ -Id \end{pmatrix} \mathbf{x} \leq \mathbf{0}, \mathbf{c}^t \mathbf{x} > 0.$$

Si ahora aplicamos el Lema de Farkas obtenemos que o bien el sistema anterior tiene solución, o bien el sistema

$$(A^t, -Id) \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \mathbf{c}, \mathbf{y} \geq \mathbf{0}, \mathbf{z} \geq \mathbf{0}.$$

Pero ahora esto puede escribirse como

$$A^t \mathbf{y} - \mathbf{z} = \mathbf{c}, \mathbf{y} \geq \mathbf{0}, \mathbf{z} \geq \mathbf{0}.$$

Y como $\mathbf{z} \geq \mathbf{0}$, esto es equivalente a

$$A^t \mathbf{y} \geq \mathbf{c}, \mathbf{y} \geq \mathbf{0}.$$

Esto demuestra el resultado. ■

Corolario 2. Sea A una matriz $m \times n$ y sea $\mathbf{c} \in \mathbb{R}^n$. Entonces uno y sólo uno de los siguientes sistemas tiene solución:

- $\exists \mathbf{x} \in \mathbb{R}^n$ tal que $A\mathbf{x} \geq \mathbf{0}$, $\mathbf{c}^t \mathbf{x} > 0$, $\mathbf{x} \geq \mathbf{0}$.
- $\exists \mathbf{y} \in \mathbb{R}^m$ tal que $A^t \mathbf{y} \geq \mathbf{c}$, $\mathbf{y} \leq \mathbf{0}$.

Demostración: El primer sistema se puede escribir como

$$-A\mathbf{x} \leq \mathbf{0}, \mathbf{c}^t \mathbf{x} > 0, \mathbf{x} \geq \mathbf{0}.$$

Si ahora aplicamos el resultado anterior, obtenemos que o bien el sistema anterior tiene solución, o bien el sistema

$$-A^t \mathbf{y} \geq \mathbf{c}, \mathbf{y} \geq \mathbf{0}.$$

Pero ahora esto puede escribirse como

$$A^t \mathbf{z} \geq \mathbf{c}, \mathbf{z} \leq \mathbf{0}.$$

Esto demuestra el resultado. ■

Corolario 3. Sea A una matriz $m \times n$, sea B una matriz $l \times n$ y sea $\mathbf{c} \in \mathbb{R}^n$. Entonces uno y sólo uno de los siguientes sistemas tiene solución:

- $\exists \mathbf{x} \in \mathbb{R}^n$ tal que $A\mathbf{x} \leq \mathbf{0}$, $B\mathbf{x} = \mathbf{0}$, $\mathbf{c}^t \mathbf{x} > 0$.
- $\exists \mathbf{y} \in \mathbb{R}^m, \exists \mathbf{z} \in \mathbb{R}^l$ tales que $A^t \mathbf{y} + B^t \mathbf{z} = \mathbf{c}$, $\mathbf{y} \geq \mathbf{0}$.

Demostración: El primer sistema se puede escribir como

$$\begin{pmatrix} A \\ B \\ -B \end{pmatrix} \mathbf{x} \leq \mathbf{0}, \mathbf{c}^t \mathbf{x} > 0.$$

Si ahora aplicamos el Lema de Farkas obtenemos que o bien el sistema anterior tiene solución, o bien el sistema

$$(A^t, B^t, -B^t) \begin{pmatrix} \mathbf{y} \\ \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} = \mathbf{c}, \mathbf{y} \geq \mathbf{0}, \mathbf{z}_1 \geq \mathbf{0}, \mathbf{z}_2 \geq \mathbf{0}.$$

Pero ahora esto puede escribirse como

$$A^t \mathbf{y} + B^t (\mathbf{z}_1 - \mathbf{z}_2) = \mathbf{c}, \mathbf{y} \geq \mathbf{0}, \mathbf{y} \geq \mathbf{0}.$$

Y definiendo $\mathbf{z} = \mathbf{z}_1 - \mathbf{z}_2$, se tiene que

$$A^t \mathbf{y} + B^t \mathbf{z} = \mathbf{c}, \mathbf{y} \geq \mathbf{0}.$$

Esto demuestra el resultado. ■

Veamos otros Teoremas de Alternativa:

Teorema 6. (Teorema de Gordan) Sea A una matriz $m \times n$. Entonces uno y sólo uno de los siguientes sistemas tiene solución:

- $\exists \mathbf{x} \in \mathbb{R}^n$ tal que $A\mathbf{x} < \mathbf{0}$.
- $\exists \mathbf{y} \in \mathbb{R}^m$ tal que $A^t \mathbf{y} = \mathbf{0}$, $\mathbf{y} \geq \mathbf{0}, \mathbf{y} \neq \mathbf{0}$.

Demostración: Supongamos que exista un vector \mathbf{x} solución del primer sistema. Si existiese un vector \mathbf{y} solución del segundo sistema, entonces se tiene:

$$\mathbf{y}^t A\mathbf{x} = \begin{cases} \mathbf{y}^t (A\mathbf{x}) = \mathbf{y}^t \mathbf{w} < 0 \\ (\mathbf{y}^t A)\mathbf{x} = \mathbf{0}^t \mathbf{x} = 0 \end{cases}$$

Y esto nos lleva a contradicción.

Supongamos ahora que el primer sistema no tenga solución. Sean los conjuntos

$$S_1 := \{\mathbf{z} : A\mathbf{x} = \mathbf{z}, \mathbf{x} \in \mathbb{R}^n\}, S_2 := \{\mathbf{z} : \mathbf{z} < \mathbf{0}\}.$$

Por hipótesis $S_1 \cap S_2 = \emptyset$, y además ambos conjuntos son convexos y no vacíos. Entonces podemos usar los teoremas de separación para concluir que existe un vector $\mathbf{y} \neq \mathbf{0}$ tal que $\mathbf{y}^t \mathbf{A} \mathbf{x} \geq \mathbf{y}^t \mathbf{z}$, $\forall \mathbf{z} < \mathbf{0}, \forall \mathbf{x} \in \mathbb{R}^n$.

Haciendo $\mathbf{x} = \mathbf{0}$, se tiene que $0 \geq \mathbf{y}^t \mathbf{z}$, $\forall \mathbf{z} < \mathbf{0}$. Si $\mathbf{y} \not\geq \mathbf{0}$, entonces $\exists i$ tal que $y_i < 0$ y entonces para $\mathbf{z} = -\mathbf{e}_i$ se tiene que $\mathbf{y}^t(-\mathbf{e}_i) = -y_i > 0$, lo que nos lleva a contradicción. Por otra parte, haciendo que las coordenadas de \mathbf{z} se aproximen a $\mathbf{0}$ se concluye que $\mathbf{y}^t \mathbf{A} \mathbf{x} \geq 0$, $\forall \mathbf{x}$. Si ahora hacemos $\mathbf{x} = -A^t \mathbf{y}$, entonces

$$0 \leq \mathbf{y}^t \mathbf{A} \mathbf{x} = -\|\mathbf{y}^t A\|^2 \leq 0.$$

En definitiva, $\mathbf{y}^t A = \mathbf{0}$ y entonces el segundo sistema tiene solución. ■

Teorema 7. (Teorema de Gale) Sea A una matriz $m \times n$ y sea $\mathbf{b} \in \mathbb{R}^n$. Entonces uno y sólo uno de los siguientes sistemas tiene solución:

- $A \mathbf{x} = \mathbf{b}$, $\mathbf{x} \in \mathbb{R}^n$.
- $A^t \mathbf{y} = \mathbf{0}$, $\mathbf{b}^t \mathbf{y} = 1$, $\mathbf{y} \in \mathbb{R}^m$.

Demostración: Supongamos que exista un vector \mathbf{x} solución del primer sistema. Si existiese un vector \mathbf{y} solución del segundo sistema, entonces se tiene:

$$\mathbf{y}^t A \mathbf{x} = \begin{cases} \mathbf{y}^t (A \mathbf{x}) = \mathbf{y}^t \mathbf{b} = 1 \\ (\mathbf{y}^t A) \mathbf{x} = \mathbf{0}^t \mathbf{x} = 0 \end{cases}$$

Y esto nos lleva a contradicción.

Supongamos ahora que el primer sistema no tenga solución. Sea el conjunto

$$S_1 := \{\mathbf{z} : A \mathbf{x} = \mathbf{z}, \mathbf{x} \in \mathbb{R}^n\}.$$

Por hipótesis $\mathbf{b} \notin S_1$, y además S_1 es un conjunto convexo, no vacío y cerrado. Entonces podemos usar los teoremas de separación para concluir que existe un vector \mathbf{y} tal que $\mathbf{y}^t \mathbf{b} > \mathbf{y}^t A \mathbf{x}$, $\forall \mathbf{x}$.

Como \mathbf{x} puede ser cualquier vector, si $\mathbf{y}^t A \neq \mathbf{0}$, podríamos hacer $\mathbf{y}^t A \mathbf{x}$ arbitrariamente grandes sin más que tomar $\mathbf{x} = \lambda A^t \mathbf{y}$ con $\lambda > 0$, y por tanto mayores que $\mathbf{y}^t \mathbf{b}$. Luego $\mathbf{y}^t A = \mathbf{0}$; pero entonces tenemos $\mathbf{y}^t \mathbf{b} > 0$. Ahora basta tomar $\lambda \mathbf{y}$ de forma que $\mathbf{y}^t \mathbf{b} = 1$. ■

2. Condiciones de Karush-Kuhn-Tucker para Programación no Lineal

2.1. Introducción

El problema general de Programación No Lineal adopta la siguiente forma:

$$\begin{array}{ll} \text{mín} & f(\mathbf{x}) \\ \text{sujeto a} & g_i(\mathbf{x}) \leq 0, i = 1, \dots, m_1 \\ & h_j(\mathbf{x}) = 0, j = m_1 + 1, \dots, m \end{array},$$

donde $f, g_i, h_j : \mathbb{R}^n \rightarrow \mathbb{R}$, $\forall i, j$. En el caso general, tanto f como las funciones que definen el recinto de soluciones factibles g_i, h_j no son necesariamente lineales. En consecuencia,

tampoco podremos asegurar que el recinto de soluciones factibles sea un conjunto convexo; como esta propiedad era básica en el método simplex, tendremos que buscar otra alternativa de resolución.

Debe tenerse en cuenta que tenemos una situación muy general, con lo que los resultados serán también muy generales. Para situaciones más concretas de problemas de Programación No Lineal existen algoritmos que resuelven estos problemas de forma eficiente. Esto da lugar a distintos problemas tipo dentro de la Programación No Lineal, como la Programación Cuadrática, Programación Separable, Programación Convexa, etc.

Una primera diferencia respecto a los problemas de Programación Lineal es que la solución óptima no es necesariamente un punto extremo del conjunto de soluciones factibles, incluso aunque este recinto sea un politopo convexo.

Ejemplo 2. *Consideremos el problema*

$$\begin{array}{ll} \text{mín} & f(x) = (x - 2)^2 + (y - 3)^2 \\ \text{sujeto a} & \begin{array}{ll} 2x + 3y & \leq 15 \\ -2x + y & \leq 1 \\ x \geq 0, y \geq 0 \end{array} \end{array}$$

En este caso las líneas de nivel son circunferencias de centro $(2, 3)$, por lo que este punto es el óptimo de la función. Sin embargo, este punto está en el interior de la región factible, que es un politopo.

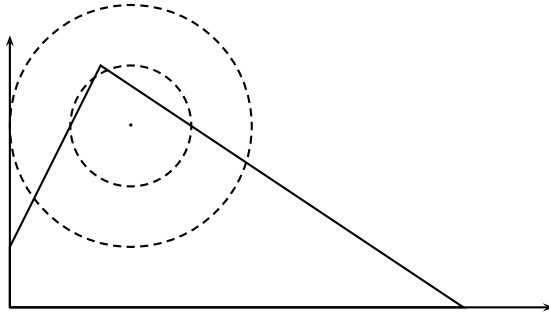


Figura 6: Gráfica del Ejemplo 2.

En un problema de Programación Lineal se podía llegar a tres tipos de final al aplicar el método simplex: solución óptima finita (única, rayo de soluciones, solución múltiple), solución no acotada, problema infactible. En el caso general de Programación No Lineal, tenemos una nueva posibilidad: función objetivo acotada pero sin solución óptima. Esto es lo que ocurre en el siguiente ejemplo.

Ejemplo 3. *Consideremos el problema*

$$\begin{array}{ll} \text{mín} & f(x) = \frac{1}{x} \\ \text{sujeto a} & x \geq 1 \end{array}$$

En este caso $f(x) \geq 0$ para cualquier x factible, pero ningún x es mínimo de la función en la región factible, puesto que $\lim_{x \rightarrow \infty} f(x) = 0$, y este límite nunca se alcanza.

2.2. Problema sin restricciones

Vamos entonces a comenzar con el planteamiento del problema. Consideremos en primer lugar con el problema sin restricciones, es decir, el problema

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{sujeto a} & \mathbf{x} \in \mathbb{R}^n \end{array}.$$

Dada la función objetivo f estamos buscando un **mínimo global**, es decir, un punto \mathbf{x}_0 tal que

$$f(\mathbf{x}_0) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n.$$

Para hallar este mínimo global nosotros vamos a buscar **mínimos locales**, es decir, puntos que optimizan la función f en un entorno suyo. Así, \mathbf{x}_0 será un mínimo local si verifica que existe un $\delta > 0$ tal que

$$f(\mathbf{x}_0) \leq f(\mathbf{x}), \forall \mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\| \leq \delta.$$

Si el problema tiene una solución óptima finita, una vez obtenidos todos los mínimos locales se trataría de comparar los valores de las funciones en esos puntos para obtener el mínimo global. Nótese sin embargo que es posible que ninguno de estos mínimos locales sea el mínimo global.

Si un punto \mathbf{x}_0 es un mínimo local, entonces es menor o igual que cualquier otro punto en un entorno. Esto implica que en cualquier dirección en la que nos movamos no mejoraremos el valor de la función en un entorno. Es decir, si \mathbf{x}_0 es un mínimo local, se cumple que para cualquier dirección $\mathbf{d} \neq \mathbf{0}$, existe un $\delta > 0$ tal que

$$f(\mathbf{x}_0) \leq f(\mathbf{x}_0 + \lambda \mathbf{d}), \forall \lambda \in [0, \delta].$$

Si una dirección \mathbf{d} no verifica la condición anterior, entonces siguiendo esa dirección obtendremos puntos en los que disminuye el valor de la función f . Diremos entonces que \mathbf{d} es una **dirección de descenso**. En definitiva, si \mathbf{x}_0 es un mínimo local, entonces no tiene direcciones de descenso; ésta será la condición que nosotros usaremos para detectar los puntos candidatos a mínimo global. Denotaremos por $\mathbf{D}(\mathbf{x}_0)$ al conjunto de direcciones de descenso para el punto \mathbf{x}_0 .

El siguiente problema es cómo hallar las direcciones de descenso en un punto \mathbf{x}_0 dado. Si f es diferenciable, entonces se tiene el siguiente resultado:

Proposición 1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable en \mathbf{x}_0 . Si existe un vector $\mathbf{d} \in \mathbb{R}^n$ tal que $\nabla f(\mathbf{x}_0)^t \mathbf{d} < 0$, entonces \mathbf{d} es una dirección de descenso de f en \mathbf{x}_0 .

Demostración: Si consideramos un desarrollo de Taylor alrededor de \mathbf{x}_0 se tiene que

$$f(\mathbf{x}_0 + \lambda \mathbf{d}) = f(\mathbf{x}_0) + \lambda \nabla f(\mathbf{x}_0)^t \mathbf{d} + o(\lambda \|\mathbf{d}\|), \forall \lambda > 0.$$

Entonces, dividiendo entre λ se tiene que

$$\frac{f(\mathbf{x}_0 + \lambda \mathbf{d}) - f(\mathbf{x}_0)}{\lambda} = \nabla f(\mathbf{x}_0)^t \mathbf{d} + \frac{o(\lambda \|\mathbf{d}\|)}{\lambda}.$$

Como $\frac{o(\lambda\|\mathbf{d}\|)}{\lambda} \rightarrow 0$ cuando $\lambda \rightarrow 0$, se tiene el resultado. ■

Nota 1. Supongamos que $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$. Entonces, tomemos \mathbf{d} tal que $\nabla f(\mathbf{x}_0)^t \mathbf{d} \neq 0$; si $\nabla f(\mathbf{x}_0)^t \mathbf{d} < 0$, entonces ya hemos encontrado una dirección de descenso; en caso contrario, $\nabla f(\mathbf{x}_0)^t \mathbf{d} > 0$ y entonces $-\mathbf{d}$ sería dirección de descenso. Además, siempre es posible encontrar una dirección en estas condiciones; por ejemplo, $-\nabla f(\mathbf{x}_0)$ siempre es una dirección de descenso si $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$. Por lo tanto, la única forma de no encontrar una dirección \mathbf{d} tal que $\nabla f(\mathbf{x}_0)^t \mathbf{d} < 0$ es que $\nabla f(\mathbf{x}_0) = \mathbf{0}$. Aunque este resultado ya era conocido, lo que es interesante del desarrollo anterior es la idea de buscar direcciones de descenso. Esto será de gran utilidad cuando estudiemos el problema con restricciones.

Nótese que la condición no es necesaria. Por ejemplo, $\nabla f(\mathbf{x}_0) = \mathbf{0}$ también lo cumplen máximos y puntos de silla. Esto es debido a que el resultado no da información sobre lo que pasa si $\nabla f(\mathbf{x}_0)^t \mathbf{d} = 0$, ya que en este caso es el término $o(\lambda\|\mathbf{d}\|)$ (que desconocemos) el que determina si se mejora o no; en este caso la dirección podría ser de ascenso (es lo que ocurre con los mínimos locales, de descenso (con los máximos locales) o incluso que representase que la función es constante en esa dirección. Este será un problema que arrastraremos durante todo el tema.

En lo que sigue supondremos f diferenciable. Usaremos la notación

$$\mathbf{F}(\mathbf{x}_0) := \{\mathbf{d} \neq \mathbf{0} : \nabla f(\mathbf{x}_0)^t \mathbf{d} < 0\}.$$

Por lo visto anteriormente, se tiene $\mathbf{F}(\mathbf{x}_0) \subseteq \mathbf{D}(\mathbf{x}_0)$, pudiendo ser el contenido estricto (es lo que pasaría con un máximo local).

Como $\nabla f(\mathbf{x}_0)^t \mathbf{d} = \|\nabla f(\mathbf{x}_0)\| \|\mathbf{d}\| \cos(\nabla f(\mathbf{x}_0), \mathbf{d})$, se tiene que si $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$, entonces $\mathbf{F}(\mathbf{x}_0)$ es un hiperplano formado por las direcciones \mathbf{d} que forman un ángulo entre 90 y 270 grados con $\nabla f(\mathbf{x}_0)$.

2.3. El problema con restricciones de desigualdad

Pasemos ahora al caso general con restricciones de desigualdad. El problema que tenemos ahora es

$$\begin{array}{ll} \text{mín} & f(\mathbf{x}) \\ \text{sujeto a} & g_i(\mathbf{x}) \leq 0, i = 1, \dots, m \end{array}$$

Dado un punto factible, en este caso es posible que si nos movemos en una determinada dirección perdamos la factibilidad; esto hace que puedan existir puntos que sean mínimos locales *factibles* aunque tengan direcciones de descenso. Dado un punto factible \mathbf{x}_0 , tendremos un conjunto de **direcciones factibles**, que serán direcciones en las que no se pierde la factibilidad en un entorno; denotaremos el conjunto de direcciones factibles para un punto factible \mathbf{x}_0 por $\mathbf{E}(\mathbf{x}_0)$. Si denotamos por \mathcal{R} la región factible,

$$\mathbf{E}(\mathbf{x}_0) := \{\mathbf{d} \neq \mathbf{0} : \exists \delta_d > 0 \mid \mathbf{x}_0 + \lambda \mathbf{d} \in \mathcal{R}, \forall \lambda \in [0, \delta_d]\}.$$

Tenemos entonces que distinguir dos casos:

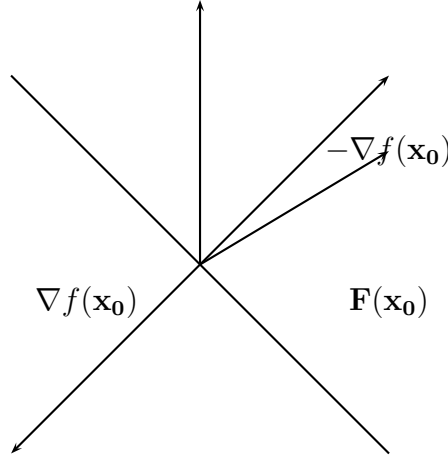


Figura 7: Interpretación gráfica del conjunto $\mathbf{F}(\mathbf{x}_0)$.

- Si \mathbf{x}_0 es un punto interior de \mathcal{R} , entonces $\mathbf{E}(\mathbf{x}_0) = \mathbb{R}^n \setminus \{\mathbf{0}\}$. Es decir, nos podemos mover en cualquier dirección sin perder la factibilidad. Nótese que si las funciones g_i son continuas, entonces \mathbf{x}_0 será un punto interior si $g_i(\mathbf{x}_0) < 0$, $i = 1, \dots, m$.
- Supongamos ahora que \mathbf{x}_0 está en la frontera de la región factible. Esto implica que, si las funciones g_i son continuas, se verifica con igualdad una o varias de las restricciones que definen \mathcal{R} . Estas restricciones se dicen restricciones **activas** y denotaremos por $I(\mathbf{x}_0)$ el conjunto de índices de las restricciones activas en \mathbf{x}_0 . Las otras restricciones no dan problemas a efectos de factibilidad y se llaman restricciones *inactivas*.

En lo que sigue supondremos que todas las funciones g_i son continuas. Sea g_i una de las restricciones activas. Entonces $g_i(\mathbf{x}_0) = 0$ y para no perder la factibilidad al seguir una dirección \mathbf{d} necesitamos que esta dirección sea una dirección de descenso o que deje la función constante. Si además g_i es diferenciable, entonces podemos aplicar la Proposición 1 y concluir que si existe un vector $\mathbf{d} \in \mathbb{R}^n$ tal que $\nabla g_i(\mathbf{x}_0)^t \mathbf{d} < 0$, entonces \mathbf{d} es una dirección de descenso de g en \mathbf{x}_0 . Usaremos la notación

$$\mathbf{G}_i(\mathbf{x}_0) := \{\mathbf{d} \neq \mathbf{0} : \nabla g_i(\mathbf{x}_0)^t \mathbf{d} < 0\}.$$

Entonces, el conjunto

$$\mathbf{G}(\mathbf{x}_0) := \bigcap_{i \in I(\mathbf{x}_0)} \mathbf{G}_i(\mathbf{x}_0)$$

nos da un conjunto de direcciones factibles para el punto \mathbf{x}_0 . Como sucedía en el caso anterior, se tiene que $\mathbf{G}(\mathbf{x}_0) \subseteq \mathbf{E}(\mathbf{x}_0)$ y en general este contenido es estricto.

Ahora, una vez determinadas las direcciones factibles para el punto \mathbf{x}_0 , tenemos que estudiar si alguna de estas direcciones es a su vez una dirección de descenso para f en \mathbf{x}_0 , es decir, estudiar el conjunto $\mathbf{D}(\mathbf{x}_0) \cap \mathbf{E}(\mathbf{x}_0)$. Si tenemos un mínimo, este conjunto es vacío.

Como en general es difícil hallar estos dos conjuntos nosotros hallaremos el conjunto $\mathbf{F}(\mathbf{x}_0) \cap \mathbf{G}(\mathbf{x}_0)$, ya que conocemos la estructura de ambos conjuntos.

En esto se basan los algoritmos de búsqueda de mínimos locales. No obstante, los puntos para los que se verifique $\mathbf{F}(\mathbf{x}_0) \cap \mathbf{G}(\mathbf{x}_0) = \emptyset$ no son necesariamente mínimos locales. Por ejemplo, si $\nabla f(\mathbf{x}_0) = \mathbf{0}$, entonces $\mathbf{F}(\mathbf{x}_0) = \emptyset$ y $\mathbf{F}(\mathbf{x}_0) \cap \mathbf{G}(\mathbf{x}_0) = \emptyset$, pero \mathbf{x}_0 no es necesariamente un mínimo local.

2.4. Condiciones de Fritz-John

Veamos las condiciones de Fritz-John para mínimo local:

Teorema 8. *Sea un punto factible \mathbf{x}_0 y denotemos por $I(\mathbf{x}_0)$ el conjunto de sus restricciones activas. Supongamos que*

- $f, g_i, i \in I(\mathbf{x}_0)$ son diferenciables en \mathbf{x}_0 .
- $g_i, i \notin I(\mathbf{x}_0)$ son continuas en \mathbf{x}_0 .

Entonces si \mathbf{x}_0 es un mínimo local de f , existen unos escalares $u_0, u_i, i \in I(\mathbf{x}_0)$ no negativos y no todos nulos tales que

$$u_0 \nabla f(\mathbf{x}_0) + \sum_{i \in I(\mathbf{x}_0)} u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

Si además $g_i, i \notin I(\mathbf{x}_0)$ son diferenciables en \mathbf{x}_0 , entonces existen unos escalares $u_0, u_i, i = 1, \dots, m$ no negativos y no todos nulos tales que

$$u_0 \nabla f(\mathbf{x}_0) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

$$u_i g_i(\mathbf{x}_0) = 0, i = 1, \dots, m.$$

Demostración: Sea \mathbf{x}_0 un mínimo local. Esto es equivalente a que no existe una dirección \mathbf{d} que sea simultáneamente de descenso y factible, lo que implica que no existe \mathbf{d} que cumple simultáneamente

$$\nabla f(\mathbf{x}_0)^t \mathbf{d} < 0, \nabla g_i(\mathbf{x}_0)^t \mathbf{d} < 0, \forall i \in I(\mathbf{x}_0).$$

Si se construye la matriz A formada por los vectores fila $\nabla f(\mathbf{x}_0)^t, \nabla g_i(\mathbf{x}_0)^t, i \in I(\mathbf{x}_0)$, se concluye que no hay un vector \mathbf{d} que cumpla

$$A\mathbf{d} < \mathbf{0}.$$

Aplicando el Teorema de Gordan, existe entonces $\mathbf{p}^t = (u_0, u_1, \dots, u_i)$ tal que

$$A^t \mathbf{p} = \mathbf{0}, \mathbf{p} \geq \mathbf{0}, \mathbf{p} \neq \mathbf{0}$$

y estas son precisamente las condiciones de Fritz-John.

La segunda parte se deriva directamente de la primera, ya que las condiciones $u_i g_i(\mathbf{x}_0) = 0$, $i = 1, \dots, m$, implican $u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}$ para las restricciones inactivas. ■

Las constantes $u_0, u_i, i \in I(\mathbf{x}_0)$ se llaman los **multiplicadores de Lagrange**. Por otra parte, las condiciones $u_i g_i(\mathbf{x}_0) = 0$ se llaman **condiciones de ortogonalidad**.

Los puntos que verifican

$$u_0 \nabla f(\mathbf{x}_0) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

$$u_i g_i(\mathbf{x}_0) = 0, i = 1, \dots, m,$$

se llaman **puntos de Fritz-John**. Lo que nos dice el resultado anterior es que todos los mínimos locales son puntos de Fritz-John. Sin embargo, no todos los puntos de Fritz-John son mínimos locales. Por ejemplo, los máximos locales de f que son factibles son también puntos de Fritz-John, ya que al ser $\nabla f(\mathbf{x}_0) = \mathbf{0}$ para estos puntos, basta tomar $u_0 = 1, u_i = 0, i \neq 0$.

Centrémonos en la primera parte de las condiciones de Fritz-John. Si las funciones que determinan las restricciones son continuas entonces, dado un punto factible \mathbf{x}_0 , para las restricciones inactivas todas las direcciones son factibles, por lo que estas restricciones no aportan restricciones y por eso no aparecen en la expresión. La expresión se puede poner como

$$u_0 \nabla f(\mathbf{x}_0) = \sum_{i \in I(\mathbf{x}_0)} u_i (-\nabla g_i(\mathbf{x}_0)).$$

Si $u_0 > 0$, esto significa que el gradiente de f en \mathbf{x}_0 se puede poner como combinación lineal de términos positivos de los opuestos de los gradientes de las restricciones activas. Para ver la interpretación geométrica de esto consideremos la Figura 8.

En esta figura estamos considerando sólo el caso de dos restricciones activas y dos dimensiones. Así, la condición de Fritz-John implica que $\nabla f(\mathbf{x}_0)$ está en el espacio comprendido entre $-\nabla g_1(\mathbf{x}_0)$ y $-\nabla g_2(\mathbf{x}_0)$. Ahora, el conjunto $\mathbf{F}(\mathbf{x}_0)$ es el conjunto de direcciones que forman con $-\nabla f(\mathbf{x}_0)$ un ángulo entre -90 y 90 grados; es la región que en la figura aparece en rojo. Por otra parte, el conjunto $\mathbf{G}(\mathbf{x}_0)$ es el conjunto de direcciones que forman con $-\nabla g_1(\mathbf{x}_0)$ y con $-\nabla g_2(\mathbf{x}_0)$ un ángulo entre -90 y 90 grados; es la región que en la figura aparece en amarillo. Entonces, puede verse que estas dos regiones tienen intersección vacía. Esto es debido a que la parte de $\mathbf{G}_1(\mathbf{x}_0)$ que interseca con $\mathbf{F}(\mathbf{x}_0)$ y la parte de $\mathbf{G}_2(\mathbf{x}_0)$ que interseca con $\mathbf{F}(\mathbf{x}_0)$ están a distintos lados del plano.

En este teorema se intenta recoger no sólo los puntos para los que $\mathbf{F}(\mathbf{x}_0) \cap \mathbf{G}(\mathbf{x}_0) = \emptyset$, sino también otras situaciones que no están previstas en este conjunto, en concreto los mínimos locales de f que son factibles. Así, estos puntos se consiguen haciendo $u_0 = 1$ y anulando los demás multiplicadores de Lagrange.

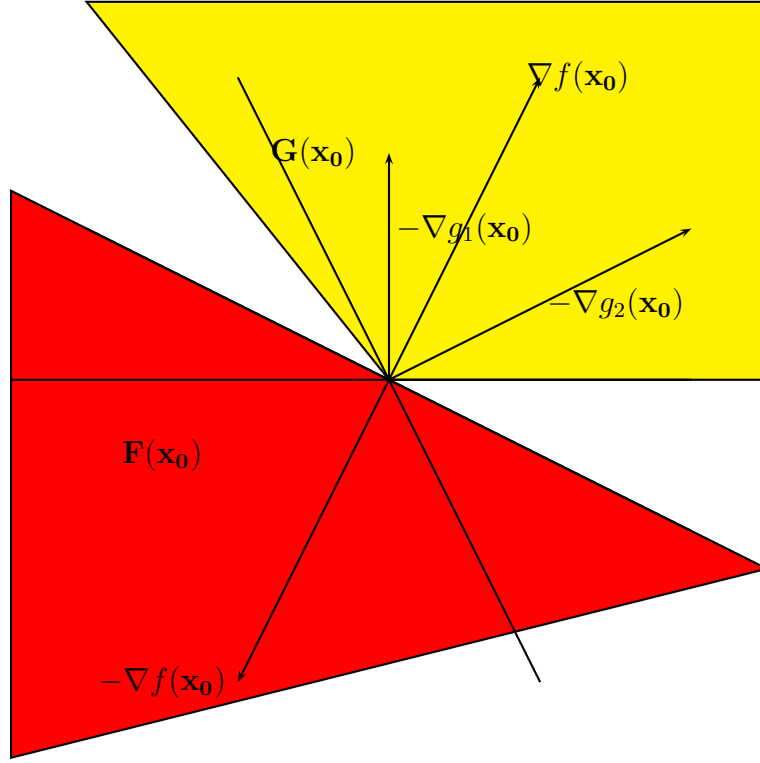


Figura 8: Interpretación gráfica de las condiciones de Fritz-John.

El problema que aparece con las condiciones de Fritz-John es que además de los mínimos locales pueden aparecer muchos otros puntos que cumplen las condiciones pero que no son mínimos locales.

2.5. Condiciones de Karush-Kuhn-Tucker

Nótese que si los vectores $\nabla g_i(\mathbf{x}_0)$, $i \in I$ no son linealmente independientes y existe una combinación lineal con coeficientes no negativos que lleve al vector $\mathbf{0}$, entonces haciendo $u_0 = 0$ también se obtiene un punto de Fritz-John, aunque en realidad este punto no sigue la filosofía en la que se basa el resultado. De hecho, \mathbf{x}_0 será un punto de Fritz-John sea cual sea la función objetivo. Esto puede dar lugar a resultados extraños como pasa en el siguiente ejemplo:

Ejemplo 4. Consideremos el problema

$$\begin{aligned} \text{mín} \quad & x \\ \text{sujeto a} \quad & y - (1 - x)^3 \leq 0 \\ & -y \leq 0 \end{aligned}$$

Consideremos el punto $(1, 0)$, que es factible. En este caso,

$$\nabla f(1, 0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad -\nabla g_1(1, 0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad -\nabla g_2(1, 0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Entonces, los vectores $-\nabla g_1(1,0)$ y $-\nabla g_2(1,0)$ son opuestos y tomando $u_0 = 0, u_1 = u_2$ se concluye que $(1,0)$ es un punto de Fritz-John. Nótese sin embargo que éste es el peor punto del dominio.

Para evitar este problema vamos a imponer que los gradientes $\nabla g_i(\mathbf{x}_0)$, $i \in I(\mathbf{x}_0)$ sean linealmente independientes. Por otra parte, siempre podemos suponer $u_0 = 1$; esto es debido a que si $\nabla f(\mathbf{x}_0)$ se anula, no importa el valor que asignemos a u_0 , pero si no se anula nos sirve para que necesariamente se tenga que poner $\nabla f(\mathbf{x}_0)$ como combinación lineal de los gradientes negativos de las restricciones activas. Esto nos lleva a las condiciones de Karush-Kuhn-Tucker.

Teorema 9. Sea un punto factible \mathbf{x}_0 y denotemos por $I(\mathbf{x}_0)$ el conjunto de sus restricciones activas. Supongamos que

- $f, g_i, i \in I(\mathbf{x}_0)$ son diferenciables en \mathbf{x}_0 .
- $g_i, i \notin I(\mathbf{x}_0)$ son continuas en \mathbf{x}_0 .
- $\nabla g_i(\mathbf{x}_0), i \in I(\mathbf{x}_0)$ son linealmente independientes.

Entonces si \mathbf{x}_0 es un mínimo local de f , existen unos escalares $u_i, i \in I(\mathbf{x}_0)$ no negativos tales que

$$\nabla f(\mathbf{x}_0) + \sum_{i \in I(\mathbf{x}_0)} u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

Si además $g_i, i \notin I(\mathbf{x}_0)$ son diferenciables en \mathbf{x}_0 , entonces existen unos escalares $u_i, i = 1, \dots, m$ no negativos tales que

$$u_0 \nabla f(\mathbf{x}_0) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

$$u_i g_i(\mathbf{x}_0) = 0, i = 1, \dots, m.$$

Demostración: Aplicando las condiciones de Fritz-John, se tiene que existen $u_0, u_i, i \in I(\mathbf{x}_0)$ no negativos y no todos nulos tales que

$$u_0 \nabla f(\mathbf{x}_0) + \sum_{i \in I(\mathbf{x}_0)} u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

Ahora, por la condición de independencia lineal u_0 no puede anularse, con lo que basta reescribir el resultado para $u_0 = 1$ y $u_i = \frac{u_i}{u_0}, i \in I(\mathbf{x}_0)$. ■

Veamos entonces el esquema de resolución para hallar los puntos Karush-Kuhn-Tucker: Construimos la función lagrangiana

$$L = f(\mathbf{x}_0) + \sum_{i=1}^m u_i g_i(\mathbf{x}_0).$$

Un punto de Karush-Kuhn-Tucker tiene que verificar las siguientes condiciones:

- Condiciones de gradiente:

$$\frac{\partial L}{\partial x_i} = 0, i = 1, \dots, n.$$

- Condiciones de ortogonalidad:

$$u_i g_i(\mathbf{x}) = 0, i = 1, \dots, m.$$

- Condiciones de factibilidad:

$$g_i(\mathbf{x}) \leq 0, i = 1, \dots, m.$$

- Condiciones de independencia:

$$\{\nabla g_i(\mathbf{x}), i \in I(\mathbf{x}_0)\} \text{ l.i.}$$

Nótese que básicamente las condiciones de Karush-Kuhn-Tucker son las condiciones de Lagrange para optimizar funciones con restricciones. La diferencia radica en que aquí estamos exigiendo la no-negatividad de los coeficientes, lo que elimina muchos de los puntos que no son mínimos locales.

Ejemplo 5. Consideremos el problema

$$\begin{array}{llll} \text{mín} & (x-2)^2 & +(y-5)^2 & \\ \text{s.a.} & (x-4)^2 & +(y-3)^2 & -4 \leq 0 \\ & -x & +y & \leq 0 \end{array}$$

La función lagrangiana es:

$$L = (x-2)^2 + (y-5)^2 + u_1 [(x-4)^2 + (y-3)^2 - 4] + u_2 [y-x].$$

Entonces,

$$\frac{\partial L}{\partial x} = 2(x-2) + 2u_1(x-4) - u_2, \quad \frac{\partial L}{\partial y} = 2(y-5) + 2u_1(y-3) + u_2.$$

De esta forma las condiciones de Karush-Kuhn-Tucker son:

$$\begin{array}{ll} 2(x-2) + 2u_1(x-4) - u_2 = 0 & (G1) \\ 2(y-5) + 2u_1(y-3) + u_2 = 0 & (G2) \\ (x-4)^2 + (y-3)^2 - 4 \leq 0 & (F1) \\ y - x \leq 0 & (F2) \\ u_1 [(x-4)^2 + (y-3)^2 - 4] = 0 & (O1) \\ u_2 [y-x] = 0 & (O2) \\ u_1, u_2 \geq 0 & (\text{no negatividad}) \end{array}$$

Ahora el problema se resuelve por casos:

■ $u_1 > 0$. Entonces, $(x-4)^2 + (y-3)^2 - 4 = 0$ (O1).

• $u_2 > 0$. Entonces $y - x = 0$ (O2). Tenemos entonces que resolver el sistema

$$\left. \begin{array}{l} (x-4)^2 + (y-3)^2 - 4 = 0 \\ y - x = 0 \end{array} \right\} \Rightarrow (x-4)^2 + (x-3)^2 - 4 = 0 \Leftrightarrow 2x^2 - 14x + 21 = 0$$

$$x = \frac{14 \pm \sqrt{196 - 168}}{4} = \frac{14 \pm \sqrt{28}}{4}.$$

Tenemos entonces los puntos $P_1 : \left(\frac{7+\sqrt{7}}{2}, \frac{7+\sqrt{7}}{2}\right)$, $P_2 : \left(\frac{7-\sqrt{7}}{2}, \frac{7-\sqrt{7}}{2}\right)$ que son factibles. Vamos a aplicar las condiciones de gradiente:

○ Para P_1 se tiene

$$\left. \begin{array}{l} 2\left(\frac{7+\sqrt{7}}{2} - 2\right) + 2u_1\left(\frac{7+\sqrt{7}}{2} - 4\right) - u_2 = 0 \\ 2\left(\frac{7+\sqrt{7}}{2} - 5\right) + 2u_1\left(\frac{7+\sqrt{7}}{2} - 3\right) + u_2 = 0 \end{array} \right\} \Rightarrow 2\sqrt{7} + 2u_1\sqrt{7} = 0 \Leftrightarrow u_1 = -1\#$$

○ Para P_2 se tiene

$$\left. \begin{array}{l} 2\left(\frac{7-\sqrt{7}}{2} - 2\right) + 2u_1\left(\frac{7-\sqrt{7}}{2} - 4\right) - u_2 = 0 \\ 2\left(\frac{7-\sqrt{7}}{2} - 5\right) + 2u_1\left(\frac{7-\sqrt{7}}{2} - 3\right) + u_2 = 0 \end{array} \right\} \Rightarrow -2\sqrt{7} - 2u_1\sqrt{7} = 0 \Leftrightarrow u_1 = -1\#$$

• $u_2 = 0$. Tenemos entonces el sistema:

$$\left. \begin{array}{l} 2(x-2) + 2u_1(x-4) = 0 \text{ (G1)} \Rightarrow x = \frac{2+4u_1}{1+u_1} \\ 2(y-5) + 2u_1(y-3) = 0 \text{ (G2)} \Rightarrow y = \frac{5+3u_1}{1+u_1} \\ (x-4)^2 + (y-3)^2 - 4 = 0 \text{ (O1)} \end{array} \right\} \Rightarrow \left(\frac{2+4u_1}{1+u_1} - 4\right)^2 + \left(\frac{5+3u_1}{1+u_1} - 3\right)^2 - 4 = 0$$

Entonces,

$$\left(\frac{-2}{1+u_1}\right)^2 + \left(\frac{2}{1+u_1}\right)^2 - 4 = 0 \Leftrightarrow \frac{2}{1+u_1} = \pm\sqrt{2}.$$

Tenemos entonces que $u_1 = \frac{2-\sqrt{2}}{\sqrt{2}}$ o bien $u_1 = \frac{\sqrt{2}-2}{\sqrt{2}}$. El segundo caso da $u_1 < 0$ y no sirve. Para el primer caso:

$$\begin{aligned} x &= \frac{2+4u_1}{1+u_1} = \frac{2+\frac{8-4\sqrt{2}}{\sqrt{2}}}{1+\frac{2-\sqrt{2}}{\sqrt{2}}} = \frac{8-2\sqrt{2}}{2} = 4 - \sqrt{2} \\ y &= \frac{5+3u_1}{1+u_1} = \frac{2+\frac{6-3\sqrt{2}}{\sqrt{2}}}{1+\frac{2-\sqrt{2}}{\sqrt{2}}} = \frac{6+2\sqrt{2}}{2} = 3 + \sqrt{2} \end{aligned}$$

Tenemos entonces el punto $P_3 : (4 - \sqrt{2}, 3 + \sqrt{2})$ pero este punto no es factible (F2).

■ $u_1 = 0$.

- $u_2 > 0$. Entonces $y - x = 0$ (O2). Tenemos entonces que resolver el sistema

$$\left. \begin{array}{rcl} 2(x-2) - u_2 & = & 0 \\ 2(y-5) + u_2 & = & 0 \\ y - x & = & 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{rcl} 2(x-2) - u_2 & = & 0 \\ 2(x-5) + u_2 & = & 0 \end{array} \right\} \Rightarrow 2(2x-7) = 0 \Leftrightarrow x = \frac{7}{2}.$$

Tenemos entonces el punto $P_4 : (\frac{7}{2}, \frac{7}{2})$. Para este punto se tiene:

$$2(x-2) - u_2 = 0 \Leftrightarrow u_2 = 3.$$

Como además es factible, entonces P_4 es un punto KKT. Nótese que sólo tiene activa una restricción (F2) y que $\nabla g_2(7/2, 7/2) = (-1, 1)^t \neq \mathbf{0}$.

- $u_2 = 0$. Tenemos entonces el sistema:

$$\left. \begin{array}{rcl} 2(x-2) = 0 & (G1) & \Leftrightarrow x = 2 \\ 2(y-5) = 0 & (G2) & \Leftrightarrow y = 5 \end{array} \right\}.$$

Tenemos entonces el punto $P_5 : (2, 5)$ pero este punto no es factible pues viola (F1) y (F2).

Veamos finalmente si $(7/2, 7/2)$ es mínimo local. Para ello veremos que $\mathbf{D}(7/2, 7/2) \cap \mathbf{E}(7/2, 7/2) = \emptyset$. Nótese que la única restricción activa de P_4 es la segunda.

Tenemos que $\nabla g_2(7/2, 7/2) = (-1, 1)^t$. Por lo tanto, tenemos que una dirección $\mathbf{d} = (d_1, d_2)^t$ es dirección de descenso si

$$\nabla g_2(7/2, 7/2)^t \mathbf{d} < 0 \Leftrightarrow -d_1 + d_2 < 0 \Leftrightarrow d_2 < d_1.$$

Y será dirección de ascenso si $d_2 > d_1$. Veamos ahora la situación si $d_2 = d_1$:

$$g_2(P_4 + \lambda \mathbf{d}) = \frac{7}{2} + \lambda d_2 - \frac{7}{2} - \lambda d_1 = 0.$$

Luego es una dirección factible. En definitiva,

$$\mathbf{E}(7/2, 7/2) = \{\mathbf{d} \neq \mathbf{0} : d_2 \leq d_1\}.$$

Ahora tenemos que $\nabla f(7/2, 7/2) = (3, -3)^t$. Por lo tanto, tenemos que una dirección $\mathbf{d} = (d_1, d_2)^t$ es dirección de descenso si

$$\nabla f(7/2, 7/2)^t \mathbf{d} < 0 \Leftrightarrow d_1 - d_2 < 0 \Leftrightarrow d_1 < d_2.$$

Y será dirección de ascenso si $d_1 > d_2$. Veamos ahora la situación si $d_2 = d_1$:

$$f(P_4 + \lambda \mathbf{d}) = \left(\frac{3}{2} + \lambda d_1\right)^2 - \left(-\frac{3}{2} + \lambda d_2\right)^2 = f(P_4) + \lambda^2 d_1^2 > f(P_4).$$

Luego no es una dirección de descenso. En definitiva,

$$\mathbf{D}(7/2, 7/2) = \{\mathbf{d} \neq \mathbf{0} : d_2 > d_1\}.$$

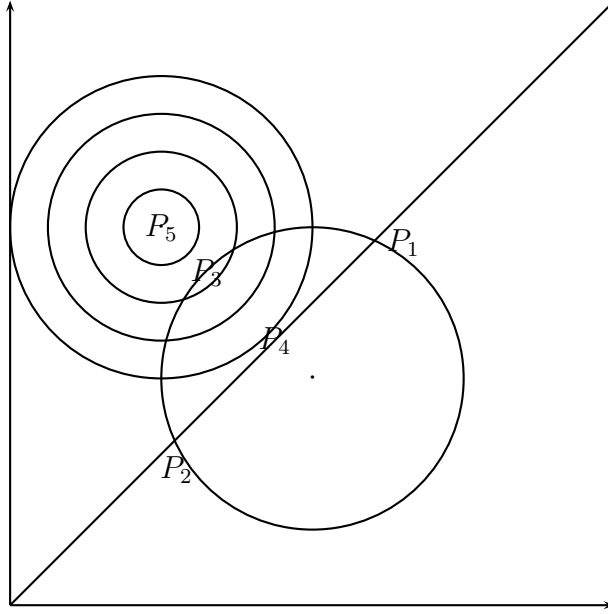


Figura 9: Gráfica del Ejemplo 5.

Con los puntos de Karush-Kuhn-Tucker se limita el número de posibles candidatos a mínimo de la función en la región factible, lo que hace que estas condiciones sean preferibles a las de Fritz-John. Sin embargo, todavía pueden aparecer puntos que no sean mínimos locales, pues no se evita la situación de puntos en los que se anula el gradiente de f ; además, es posible que algún mínimo local no cumpla las condiciones de Karush-Kuhn-Tucker debido a la tercera condición.

2.6. El problema con restricciones de desigualdad y de igualdad

Pasemos ahora al caso general con restricciones de desigualdad y de igualdad. El problema que tenemos ahora es

$$\begin{array}{ll} \text{mín} & f(\mathbf{x}) \\ \text{sujeto a} & g_i(\mathbf{x}) \leq 0, i = 1, \dots, m_1 \\ & h_i(\mathbf{x}) = 0, i = m_1 + 1, \dots, m \end{array}$$

Nótese que ahora las restricciones de igualdad son siempre restricciones activas. Por otra parte, para tratar este problema vamos a aprovechar el esquema desarrollado para el caso de tener sólo restricciones de desigualdad. Para ello, transformaremos las restricciones $h_i(\mathbf{x}) = 0$ en

$$h_i(\mathbf{x}) = 0 \Leftrightarrow \begin{cases} h_i(\mathbf{x}) \leq 0 \\ h_i(\mathbf{x}) \geq 0 \Leftrightarrow -h_i(\mathbf{x}) \leq 0 \end{cases}$$

De esta forma, la función de los multiplicadores de Lagrange pasaría a ser

$$L = f(\mathbf{x}_0) + \sum_{i=1}^{m_1} u_i g_i(\mathbf{x}_0) + \sum_{i=m_1}^m v_i^1 h_i(\mathbf{x}_0) - \sum_{i=m_1}^m v_i^2 h_i(\mathbf{x}_0),$$

donde $u_i \geq 0, i = 1, \dots, m_1, v_i^1, v_i^2 \geq 0, i = m_1, \dots, m$. Y haciendo $v_i := v_i^1 - v_i^2$ se concluye

$$L = f(\mathbf{x}_0) + \sum_{i=1}^{m_1} u_i g_i(\mathbf{x}_0) + \sum_{i=m_1}^m v_i h_i(\mathbf{x}_0), u_i \geq 0, i = 1, \dots, m_1, v_i \in \mathbb{R}, i = m_1, \dots, m.$$

En definitiva se tienen los siguientes resultados:

Teorema 10. Condiciones de Fritz-John. *Sea un punto factible \mathbf{x}_0 y denotemos por $I(\mathbf{x}_0)$ el conjunto de sus restricciones de desigualdad activas. Supongamos que*

- $f, g_i, i \in I(\mathbf{x}_0), h_j, j = m_1, \dots, m$, son diferenciables en \mathbf{x}_0 .
- $g_i, i \notin I(\mathbf{x}_0)$ son continuas en \mathbf{x}_0 .

Entonces si \mathbf{x}_0 es un mínimo local de f , existen unos escalares no todos nulos y tales que $u_0, u_i, i \in I(\mathbf{x}_0)$ no negativos, $v_i, i = m_1, \dots, m$ reales, tales que

$$u_0 \nabla f(\mathbf{x}_0) + \sum_{i \in I(\mathbf{x}_0)} u_i \nabla g_i(\mathbf{x}_0) + \sum_{i=m_1}^m v_i \nabla h_i(\mathbf{x}_0) = \mathbf{0}.$$

Si además $g_i, i \notin I(\mathbf{x}_0)$ son diferenciables en \mathbf{x}_0 , entonces existen unos escalares no todos nulos y tales que $u_0, u_i, i \in I(\mathbf{x}_0)$ no negativos, $v_i, i = m_1, \dots, m$ reales tales que

$$u_0 \nabla f(\mathbf{x}_0) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}_0) + \sum_{i=m_1}^m v_i \nabla h_i(\mathbf{x}_0) = \mathbf{0}.$$

$$u_i g_i(\mathbf{x}_0) = 0, i = 1, \dots, m.$$

Demostración: Como en el caso de restricciones de desigualdad, basta demostrar la primera parte. Para demostrar este resultado necesitamos usar el siguiente teorema:

Teorema 11. *Sea un punto factible \mathbf{x}_0 y denotemos por $I(\mathbf{x}_0)$ el conjunto de sus restricciones de desigualdad activas. Supongamos que*

- $f, g_i, i \in I(\mathbf{x}_0), h_j, j = m_1, \dots, m$, son diferenciables en \mathbf{x}_0 .
- $g_i, i \notin I(\mathbf{x}_0)$ son continuas en \mathbf{x}_0 .
- $\{\nabla h_i(\mathbf{x}_0)\}_{i=1, \dots, m}$ son linealmente independientes.

Sea

$$H(\mathbf{x}_0) := \{\mathbf{d} : \nabla h_i(\mathbf{x}_0)^t \mathbf{d} = 0, i = 1, \dots, m\}.$$

Entonces si \mathbf{x}_0 es un mínimo local de f , entonces

$$F(\mathbf{x}_0) \cup G(\mathbf{x}_0) \cup H(\mathbf{x}_0) = \emptyset.$$

Veamos entonces la demostración de las condiciones de Fritz-John. Si $\{\nabla h_i(\mathbf{x}_0)\}_{i=1,\dots,m}$ son linealmente dependientes, entonces basta hacer $u_0 = 0, u_i = 0, i \in I(\mathbf{x}_0)$ y ya se tiene el resultado.

Supongamos entonces que $\{\nabla h_i(\mathbf{x}_0)\}_{i=1,\dots,m}$ son linealmente independientes. Sea A la matriz definida por

$$A_1 := \begin{pmatrix} \nabla f(\mathbf{x}_0)^t \\ \nabla g_i(\mathbf{x}_0), i \in I(\mathbf{x}_0) \end{pmatrix}, \quad A_2 := (\nabla h_i(\mathbf{x}_0), i = 1, \dots, m).$$

Consideremos entonces

$$C_1 := \{(\mathbf{z}_1^t, \mathbf{z}_2^t) : \mathbf{z}_1 = A_1 \mathbf{d}, \mathbf{z}_2 = A_2 \mathbf{d}\}, \quad C_2 := \{(\mathbf{z}_1^t, \mathbf{0}^t) : \mathbf{z}_1 < \mathbf{0}\}.$$

Entonces C_1 y C_2 son conjuntos convexos y no vacíos. Por el resultado anterior, $C_1 \cap C_2 = \emptyset$. Entonces podemos aplicar los Teoremas de Separación de conjuntos convexos y concluir que existe $\mathbf{p}^t(\mathbf{p}_1^t, \mathbf{p}_2^t) = (u_0, u_1, i \in I(\mathbf{x}_0), v_j, j = 1, \dots, m)$ tal que

$$\mathbf{p}_1^t A_1 \mathbf{d} + \mathbf{p}_2^t A_2 \mathbf{d} \geq \mathbf{p}_1^t \mathbf{z}_1.$$

Ahora bien, como \mathbf{z}_1 puede hacerse indefinidamente grande y negativo, esto fuerza que $\mathbf{p}_1 \geq \mathbf{0}$. Por otra parte, haciendo \mathbf{z}_1 tender a $\mathbf{0}$, se tiene que

$$0 \leq \mathbf{p}_1^t A_1 \mathbf{d} + \mathbf{p}_2^t A_2 \mathbf{d} = (\mathbf{p}_1^t A_1 + \mathbf{p}_2^t A_2) \mathbf{d}.$$

Como \mathbf{d} puede ser cualquier vector no nulo, si $(\mathbf{p}_1^t A_1 + \mathbf{p}_2^t A_2) \neq \mathbf{0}$, entonces tomando $\mathbf{d}^t = (\mathbf{p}_1^t A_1 + \mathbf{p}_2^t A_2)$ se llegaría a

$$0 \leq \mathbf{p}_1^t A_1 \mathbf{d} + \mathbf{p}_2^t A_2 \mathbf{d} = -\|\mathbf{p}_1^t A_1 + \mathbf{p}_2^t A_2\|^2 \leq 0.$$

Por lo tanto, ha de ser $\mathbf{p}_1^t A_1 + \mathbf{p}_2^t A_2 = \mathbf{0}$, que es lo que nos dicen las condiciones de Fritz-John. ■

Teorema 12. Condiciones de Karush-Kuhn-Tucker. *Sea un punto factible \mathbf{x}_0 y denotemos por $I(\mathbf{x}_0)$ el conjunto de sus restricciones de desigualdad activas. Supongamos que*

- $f, g_i, i \in I(\mathbf{x}_0), h_j, j = m_1, \dots, m$ son diferenciables en \mathbf{x}_0 .
- $g_i, i \notin I(\mathbf{x}_0)$ son continuas en \mathbf{x}_0 .
- $\nabla g_i(\mathbf{x}_0), i \in I(\mathbf{x}_0), \nabla h_i(\mathbf{x}_0), i = m_1, \dots, m$, son linealmente independientes.

Entonces si \mathbf{x}_0 es un mínimo local de f , existen unos escalares $u_i, i \in I(\mathbf{x}_0)$ no negativos y unos escalares reales $v_i, i = m_1, \dots, m$ tales que

$$\nabla f(\mathbf{x}_0) + \sum_{i \in I(\mathbf{x}_0)} u_i \nabla g_i(\mathbf{x}_0) + \sum_{i=m_1}^m v_i \nabla h_i(\mathbf{x}_0) = \mathbf{0}.$$

Si además $g_i, i \notin I(\mathbf{x}_0)$ son diferenciables en \mathbf{x}_0 , entonces existen unos escalares $u_i, i = 1, \dots, m$ no negativos y unos escalares reales $v_i, i = m_1, \dots, m$ tales que

$$u_0 \nabla f(\mathbf{x}_0) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}_0) + \sum_{i=m_1}^m v_i \nabla h_i(\mathbf{x}_0) = \mathbf{0}.$$

$$u_i g_i(\mathbf{x}_0) = 0, i = 1, \dots, m.$$

Veamos entonces el esquema de resolución para hallar los puntos Karush-Kuhn-Tucker: Construimos la función lagrangiana

$$L = f(\mathbf{x}_0) + \sum_{i=1}^m u_i g_i(\mathbf{x}_0) + \sum_{i=m_1}^m v_i h_i(\mathbf{x}_0) = \mathbf{0}.$$

Un punto de Karush-Kuhn-Tucker tiene que verificar las siguientes condiciones:

- Condiciones de gradiente:

$$\frac{\partial L}{\partial x_i} = 0, i = 1, \dots, n.$$

- Condiciones de ortogonalidad:

$$u_i g_i(\mathbf{x}) = 0, i = 1, \dots, m_1.$$

- Condiciones de factibilidad:

$$g_i(\mathbf{x}) \leq 0, i = 1, \dots, m_1.$$

$$h_i(\mathbf{x}) = 0, i = m_1, \dots, m.$$

- Condiciones de independencia:

$$\nabla g_i(\mathbf{x}), i \in I(\mathbf{x}_0), \nabla h_i(\mathbf{x}), i = m_1, \dots, m, \text{ l.i.}$$

3. Funciones convexas

Vamos a estudiar en esta sección los resultados básicos de funciones convexas. Veremos las implicaciones que estas funciones tienen en el caso de los puntos KKT.

Definition 3. Dada una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, diremos que es **convexa** si para cualesquiera $\mathbf{x}_1, \mathbf{x}_2$ se cumple

$$f(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1 - \lambda) f(\mathbf{x}_2), \forall \lambda \in [0, 1].$$

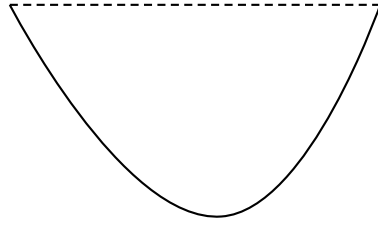


Figura 10: Ejemplo de función convexa.

En definitiva, una función convexa nos dice que la interpolación lineal está siempre por encima de la función. Si f es convexa, la función $-f$ se dice que es *cóncava*. Nótese que una función lineal es convexa trivialmente, aunque no parece a priori que la idea de función convexa esté pensada para este tipo de funciones. Por ello, se define lo siguiente:

Definition 4. Dada una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, diremos que es **estrictamente convexa** si para cualesquiera $\mathbf{x}_1, \mathbf{x}_2$ distintos, se cumple

$$f(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) < \lambda f(\mathbf{x}_1) + (1 - \lambda) f(\mathbf{x}_2), \forall \lambda \in (0, 1).$$

Se tienen las siguientes propiedades:

Lema 4. Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es convexa si y sólo si para cualesquiera $\mathbf{x}_1, \dots, \mathbf{x}_k$ se cumple

$$f(\lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k) \leq \lambda_1 f(\mathbf{x}_1) + \dots + \lambda_k f(\mathbf{x}_k), \forall \lambda_i \geq 0, i = 1, \dots, k, \sum_{i=1}^k \lambda_i = 1.$$

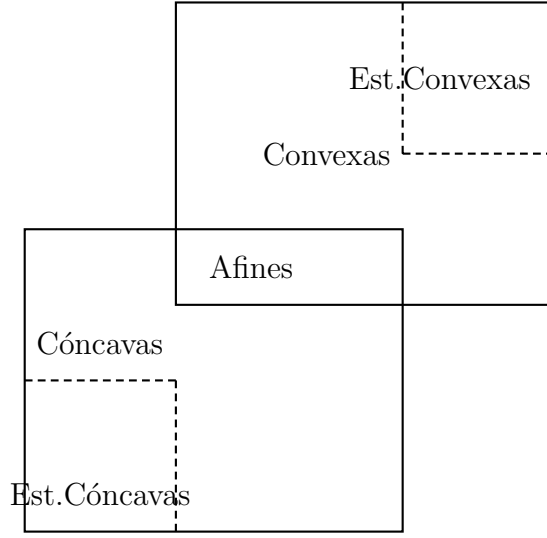


Figura 11: Relaciones entre los distintos tipos de funciones.

Demostración: \Rightarrow) La demostración se hace por inducción en k . Para $k = 2$ el resultado es cierto por hipótesis.

Sea $k > 2$ y supongamos el resultado cierto hasta $k - 1$. Entonces,

$$\lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k = \lambda_1 \mathbf{x}_1 + (1 - \lambda_1) \left[\frac{\lambda_2}{1 - \lambda_1} \mathbf{x}_2 + \dots + \frac{\lambda_k}{1 - \lambda_1} \mathbf{x}_k \right].$$

Aplicando el caso $k = 2$ e inducción se obtiene

$$\begin{aligned} f(\lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k) &= f \left(\lambda_1 \mathbf{x}_1 + (1 - \lambda_1) \left[\frac{\lambda_2}{1 - \lambda_1} \mathbf{x}_2 + \dots + \frac{\lambda_k}{1 - \lambda_1} \mathbf{x}_k \right] \right) \\ &\leq \lambda_1 f(\mathbf{x}_1) + (1 - \lambda_1) f \left(\frac{\lambda_2}{1 - \lambda_1} \mathbf{x}_2 + \dots + \frac{\lambda_k}{1 - \lambda_1} \mathbf{x}_k \right) \\ &\leq \lambda_1 f(\mathbf{x}_1) + (1 - \lambda_1) \left[\frac{\lambda_2}{1 - \lambda_1} f(\mathbf{x}_2) + \dots + \frac{\lambda_k}{1 - \lambda_1} f(\mathbf{x}_k) \right] \\ &= \lambda_1 f(\mathbf{x}_1) + \dots + \lambda_k f(\mathbf{x}_k) \end{aligned}$$

\Leftarrow) Trivial. ■

Lema 5. Dadas $f_1, \dots, f_k : \mathbb{R}^n \rightarrow \mathbb{R}$ convexas, entonces son convexas:

- $\sum_{i=1}^k \alpha_i f_i$, $\alpha_i > 0, i = 1, \dots, k$.
- $\max\{f_1, \dots, f_k\}$.

Demostración:

- Aplicando la definición, dados $\mathbf{x}_1, \mathbf{x}_2$ y $\lambda \in (0, 1)$

$$\begin{aligned} \left(\sum_{i=1}^k \alpha_i f_i \right) (\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) &= \sum_{i=1}^k \alpha_i f_i (\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) \\ &\leq \sum_{i=1}^k \alpha_i (\lambda f_i(\mathbf{x}_1) + (1 - \lambda) f_i(\mathbf{x}_2)) \\ &= \lambda \sum_{i=1}^k \alpha_i f_i(\mathbf{x}_1) + (1 - \lambda) \sum_{i=1}^k \alpha_i f_i(\mathbf{x}_2) \\ &= \lambda \left(\sum_{i=1}^k \alpha_i f_i \right) (\mathbf{x}_1) + (1 - \lambda) \left(\sum_{i=1}^k \alpha_i f_i \right) (\mathbf{x}_2) \end{aligned}$$

- Aplicando la definición, dados $\mathbf{x}_1, \mathbf{x}_2$ y $\lambda \in (0, 1)$

$$\begin{aligned} \max\{f_1, \dots, f_k\}(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2) &= \max\{f_1(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2), \dots, f_k(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2)\} \\ &\leq \max\{\lambda f_1(\mathbf{x}_1) + (1 - \lambda) f_1(\mathbf{x}_2), \dots, \lambda f_k(\mathbf{x}_1) + (1 - \lambda) f_k(\mathbf{x}_2)\} \\ &\leq \lambda \max\{f_1(\mathbf{x}_1), \dots, f_k(\mathbf{x}_1)\} + (1 - \lambda) \max\{f_1(\mathbf{x}_2), \dots, f_k(\mathbf{x}_2)\} \end{aligned}$$

■

Supongamos ahora que f es diferenciable. Entonces se tienen los siguientes resultados:

Teorema 13. *Sea una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable y sean $\mathbf{x}_1, \mathbf{x}_2$ distintos. Entonces:*

- f es convexa si y sólo si $f(\mathbf{x}_2) - f(\mathbf{x}_1) \geq \nabla f(\mathbf{x}_1)^t (\mathbf{x}_2 - \mathbf{x}_1)$.
- f es estrictamente convexa si y sólo si $f(\mathbf{x}_2) - f(\mathbf{x}_1) > \nabla f(\mathbf{x}_1)^t (\mathbf{x}_2 - \mathbf{x}_1)$.

Teorema 14. *Sea una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable y sean $\mathbf{x}_1, \mathbf{x}_2$ distintos. Entonces:*

- f es convexa si y sólo si $(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))^t (\mathbf{x}_2 - \mathbf{x}_1) \geq 0$.
- f es estrictamente convexa si y sólo si $(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))^t (\mathbf{x}_2 - \mathbf{x}_1) > 0$.

Para la última caracterización, necesitamos recordar algunos conceptos de matrices:

Definition 5. *Sea A una matriz simétrica de orden n . Entonces*

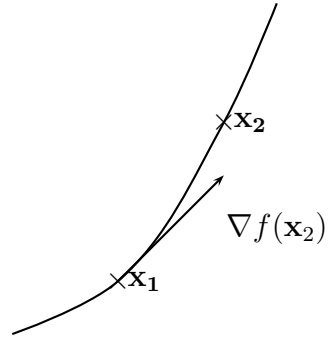


Figura 12: Interpretación gráfica del Teorema 13.

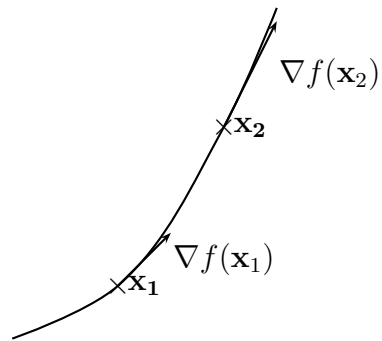


Figura 13: Interpretación gráfica del Teorema 14.

- A es **semidefinida positiva** si $\mathbf{x}^t A \mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n$.
- A es **definida positiva** si $\mathbf{x}^t A \mathbf{x} > 0, \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$.

Como para una matriz simétrica cuadrada los autovalores son números reales, podemos estudiar si la matriz es semidefinida positiva o definida positiva a partir de sus autovalores:

Proposición 2. *Sea A una matriz cuadrada de orden n con todos sus autovalores reales. Entonces:*

- A es *semidefinida positiva* si y sólo si todos sus autovalores son no negativos.
- A es *definida positiva* si y sólo si todos sus autovalores son positivos.

Se tiene ahora el siguiente resultado:

Teorema 15. *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función dos veces diferenciable y sea $H(\mathbf{x})$ la matriz hessiana en \mathbf{x} . Entonces:*

- f es *convexa* si y sólo si $H(\mathbf{x})$ es *semidefinida positiva*, $\forall \mathbf{x} \in \mathbb{R}^n$.
- Si $H(\mathbf{x})$ es *definida positiva*, $\forall \mathbf{x} \in \mathbb{R}^n$, entonces f es *estrictamente convexa*.

Las condiciones de Fritz-John y de Karush-Kuhn-Tucker nos dan condiciones para obtener posibles mínimos locales. Sin embargo, ya hemos visto que en ocasiones los puntos obtenidos con estos resultados no son mínimos locales, ni mucho menos el mínimo global. Esto es debido a que estas condiciones deben ser aplicables a cualesquiera funciones, ya sean objetivo o de restricciones. Por otra parte, hay situaciones en que estas condiciones funcionan bien. Uno de estos casos es el de las funciones convexas.

- En el caso de un problema en que sólo tenemos restricciones de desigualdad,

Teorema 16. *(Condiciones suficientes de Karush-Kuhn-Tucker). Sea \mathbf{x}^* un punto KKT, y supongamos que f y $\{g_i\}_{i \in I(\mathbf{x}^*)}$ son funciones convexas. Entonces \mathbf{x}^* es un mínimo global.*

Demostración: Como \mathbf{x}^* es un punto KKT, existen unos escalares $u_i, i \in I(\mathbf{x}_0)$ no negativos tales que

$$\nabla f(\mathbf{x}_0) + \sum_{i \in I(\mathbf{x}_0)} u_i \nabla g_i(\mathbf{x}_0) = \mathbf{0}.$$

Sea \mathbf{x} factible. Entonces,

$$g_i(\mathbf{x}) \leq 0 = g_i(\mathbf{x}^*), \forall i \in I(\mathbf{x}^*).$$

Como g_i es convexa,

$$g_i(\mathbf{x}^* + \lambda(\mathbf{x} - \mathbf{x}^*)) = g_i(\lambda \mathbf{x} + (1 - \lambda)\mathbf{x}^*) \leq \lambda g_i(\mathbf{x}) + (1 - \lambda)g_i(\mathbf{x}^*) \leq 0 = g_i(\mathbf{x}^*), \forall \lambda \in [0, 1].$$

Esto implica que $\mathbf{x} - \mathbf{x}^*$ es una dirección de no crecimiento y como g_i es diferenciable en \mathbf{x}^* ,

$$\nabla g_i(\mathbf{x}^*)^t(\mathbf{x} - \mathbf{x}^*) \leq 0, \forall i \in I(\mathbf{x}^*).$$

Si ahora multiplicamos por los valores u_i (que son no negativos) y sumamos todas estas inecuaciones, se obtiene

$$\sum_{i \in I(\mathbf{x}^*)} u_i \nabla g_i(\mathbf{x}^*)^t(\mathbf{x} - \mathbf{x}^*) \leq 0.$$

Pero por las condiciones de Karush-Kuhn-Tucker, esto obliga a que

$$\nabla f(\mathbf{x}^*)^t(\mathbf{x} - \mathbf{x}^*) \geq 0.$$

Como f es convexa,

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*)^t(\mathbf{x} - \mathbf{x}^*) \geq 0 \Rightarrow f(\mathbf{x}) \geq f(\mathbf{x}^*),$$

y \mathbf{x}^* es un mínimo local. ■

■ En el caso de un problema en que tenemos restricciones de desigualdad y de igualdad,

Teorema 17. (*Condiciones suficientes de Karush-Kuhn-Tucker*). Sea \mathbf{x} un punto KKT, y supongamos que f , $\{g_i\}_{i \in I(\mathbf{x})}$, $\{h_i\}_{i|v_i > 0}$ convexas y $\{h_i\}_{i|v_i < 0}$ cóncavas. Entonces \mathbf{x} es un mínimo global.

Por ejemplo, en los problemas de Programación Cuadrática, tenemos una función objetivo que es cuadrática, es decir,

$$f(\mathbf{x}) = \mathbf{c}^t \mathbf{x} + \frac{1}{2} \mathbf{x}^t H \mathbf{x},$$

y las restricciones son de tipo lineal. Por ello, si la función cuadrática tiene un hessiano H semidefinido positivo, un punto de KKT es solución del problema. Y los algoritmos de Programación Cuadrática, como el basado en el algoritmo de Lemke, tratan de buscar un punto KKT. Nótese que para este tipo de problemas, las condiciones de KKT son lineales.