May 9, 2023

4. OPTIMIZACIÓN CON RESTRICCIONES DE DESIGUALDAD: EL MÉTODO DE KUHN-TUCKER

A lo largo de todo el capítulo, D denota un **conjunto abierto** de $\mathbb{R}^n$.

1. Introducción

Estudiamos en este tema problemas de optimización con restricciones de desigualdad.

$$\max\ f(x) \quad \text{s.a: } x \in S, \tag{1.1}$$

donde $S = \{x \in D : \ g_1(x) \leq b_1, \ldots, g_m(x) \leq b_m\}$ y donde $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \ldots, g_m) : D \longrightarrow \mathbb{R}^m$. En contraste con los problemas de Lagrange, no existe limitación del número finito de restricciones.

Nota 1.1. Una restricción del tipo $g_i(x) \geq b_i$ es equivalente a $-g_i(x) \leq -b_i$. El problema

$$\min\ f(x) \quad \text{s.t.: } x \in S,$$

tiene las mismas soluciones que

$$\max\ -f(x) \quad \text{s.a: } x \in S.$$

2. Condiciones necesarias de Khun-Tucker

Definición 2.1. Dado un punto $x_0 \in D$, decimos que la restricción i-ésima del problema (1.1), $i = 1, 2, \ldots, m$, está **saturada** en el punto x_0, si $g_i(x_0) = b_i$. Si $g_i(p) < b_i$, entonces decimos que la restricción i-ésima no está saturada en el punto x_0 (o que x_0 no satura la restricción).

Definición 2.2. Sea g de clase C^1 en D. El punto $x_0 \in D$ es **regular** si x_0 no satura ninguan restricción, o bien los vectores gradiente de las restricciones saturadas en x_0 forman una matriz de rango máximo (es decir, en el caso de que $k \leq m$ sea el número de restricciones saturadas en el punto x_0, entonces ma matriz mencionada tiene rango k).

Definición 2.3. La función Lagrangiana asociada a (1.1) es

$$\mathcal{L}(x, \lambda) = f(x) + \lambda \cdot (b - g(x)), \tag{2.1}$$

donde $\lambda = (\lambda_1, \ldots, \lambda_m)$.

Teorema 2.4 (Método de Kuhn-Tucker)**.** Suponemos que las funciones $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \ldots, g_m) : D \longrightarrow \mathbb{R}^m$ son de clase C^1 en D y que x_0 es un punto regular de (1.1). Si x_0 es una solución del problema (1.1), entonces existe un vector de multiplicadores $\lambda_0 = (\lambda_1^0, \ldots, \lambda_m^0) \in \mathbb{R}^m$ tal que

$$\forall i \in \{1, \ldots, n\} \quad \frac{\partial \mathcal{L}}{\partial x_i}(x_0, \lambda_0) = 0,$$

$$\forall i \in \{1, \ldots, m\} \quad \left\{ \begin{array}{rcl} \lambda_i^0 (b_i - g_i(x_0)) & = & 0, \\ \lambda_i^0 & \geq & 0, \\ g_i(x_0) & \leq & b_i. \end{array} \right.$$

Nota 2.5. Las ecuaciones

(1) $\nabla_x L(x,\lambda) = 0$,
(2) $\lambda_1(b_1 - g_1(x)) = 0, \cdots, \lambda_m(b_m - g_m(x)) = 0$,
(3) $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0$,
(4) $g_1(x) \leq b_1, \ldots, g_m(x) \leq b_m$.

se llaman ecuaciones de Kuhn-Tucker del problema (1.1).

Nota 2.6. Una forma de aplicar las condiciones necesarias del Teorema 2.4 es solucionar el sistema de $n+m$ ecuaciones y $n+m$ incógnitas

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_1}(x,\lambda) & = & 0 \\ \vdots & \vdots & \vdots \\ \frac{\partial \mathcal{L}}{\partial x_n}(x,\lambda) & = & 0 \\ \lambda_1^0(b_1 - g_1(x_0)) & = & 0 \\ \vdots & \vdots & \vdots \\ \lambda_m^0(b_m - g_m(x_0)) & = & 0. \end{cases}$$

De las ecuaciones (x,λ) encontradas, sólo se retienen aquéllas que satisfacen el resto de las condiciones necesarias: $x \in S$ y $\lambda_i \geq 0$.

Ejemplo 2.7 (Substitutos perfectos). Supongamos que la renta de un agente es 5 y que su función de utilidad obre dos bienes de consumo es $u(x,y) = 2x + y$. Si los precios de los bienes son $p_1 = 3$, $p_2 = 1$ ¿cuál es la demanda de cada bien?
El problema de maximización de la utilidad del agente es

$$\begin{array}{rl} \max & 2x + y \\ \text{s.a.} & 3x + y \leq 5 \\ & x \geq 0 \\ & y \geq 0 \end{array}$$

En primer lugar escribimos este problema en la forma (1.1)

$$\begin{array}{rl} \max & 2x + y \\ \text{s.a.} & 5 - 3x - y \geq 0 \\ & x \geq 0 \\ & y \geq 0 \end{array}$$

El Lagrangiano asociado es

$$L(x,y) = 2x + y + \lambda_1(5 - 3x - y) + \lambda_2 x + \lambda_3 y$$

y las ecuaciones de Kuhn-Tucker del problema son

$$\begin{aligned}
\frac{\partial L}{\partial x} &= 2 - 3\lambda_1 + \lambda_2 = 0 \\
\frac{\partial L}{\partial y} &= 1 - \lambda_1 + \lambda_3 = 0 \\
& \lambda_1(5 - 3x - y) = 0 \\
& \lambda_2 x = 0 \\
& \lambda_3 y = 0 \\
& 3x + y \leq 5 \\
& x \geq 0 \\
& y \geq 0 \\
& \lambda_1, \lambda_2, \lambda_3 \geq 0
\end{aligned}$$

Observamos que si $\lambda_1 = 0$ entonces la primera ecuación implica que $\lambda_2 = -2 < 0$, que contradice la ecuación $\lambda_1, \lambda_2, \lambda_3 \geq 0$. Por lo tanto, $\lambda_1 > 0$. De la ecuación $\lambda_1(5 - 3x - y) = 0$ concluimos ahora que $5 - 3x - y = 0$ por lo que

$$y = 5 - 3x$$

Supongamos que $x > 0$. En este caso, la ecuación $\lambda_2 x = 0$ implica que $\lambda_2 = 0$. De la primera ecuación deducimos que $\lambda_1 = 2/3$ y sustituyendo en la segunda ecuación obtenemos que $\lambda_3 = \lambda_1 - 1 = -1/3 < 0$ que contradice la ecuación $\lambda_1, \lambda_2, \lambda_3 \geq 0$. Concluimos entonces que $x = 0$ y por tanto $y = 5$. Vemos que

$$x = 0, \quad y = 5, \quad \lambda_1 = \lambda_2 = 1, \quad \lambda_3 = 0$$

es la única solución del sistema.

Nota 2.8. Las ecuaciones Kuhn-Tucker de la proposición 2.4 sólo son válidas cuando el problema está escrito en la forma 1.1. Si, por ejemplo, se pide resolver un problema de minimización o las desigualdades son $\geq$ en lugar de $\leq$, entonces el problema se puede transformar de forma trivial para convertirlo en uno que venga descrito en la forma 1.1. Por ejemplo, si nos piden minimizar $f(x)$, esto es equivalente a maximizar $-f(x)$; o bien, si la restricción i-ésima es $g_i(x) \leq a_i$, eso equivale a la restricción $a_i - g_i(x) \geq 0$.

Nota 2.9. Si S es compacto, el Teorema de Weierstrass asegura la existencia de soluciones globales de (1.1). Asumiendo que la hipótesis de regularidad se cumple, estas soluciones globales serán puntos críticos de f relativos a S. Estas soluciones globales se determinan evaluando los puntos críticos en la función objetivo f. Los valores extremosl aon loa máximos y mínimos de f en S.

3. Condiciones suficientes. Programas convexos

Sean $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \ldots, g_m) : D \longrightarrow \mathbb{R}^m$ de clase C^1 en D.

Teorema 3.1. Si en el problema (1.1), la función objetivo f es cóncava y las restricciones $g_1, \ldots, g_m$ son convexas (recordar que las restricciones son de la forma $g_i(x) \leq b_i$), las condiciones necesarias de Khun-Tucker establecidas en el Teorema (2.4) son también suficientes y si $x_0 \in S$ las verifica, entonces x_0 es un máximo global de f en S.

Nota 3.2. Si además de las hipótesis anteriores, f es estrictamente cóncava, entonces el máximo global, de existir, es único. Por tanto, una vez que una solución (x_0, λ_0) de las condiciones de KT ha sido hallada, podemos parar la búsqueda, puesto que no habrá mś soluciones.

4. Restricciones de no negatividad

Muy a menudo los problemas con restricciones incluyen la no negatividad de las variables. A pesar de que este caso es un caso particular de la teoría desarrollada anteriormente, en muchos libros de texto se describen las condiciones de Khun-Tucker para este caso que, a primera vista, pueden parecer diferentes a las ya descritas. En realidad, las condiciones son equivalentes, y tienen la siguiente estructura. Sea el problema de optimización con restricciones de no negatividad siguiente:

$$\max\ f(x) \quad \text{s.a: } x \in S, \tag{4.1}$$

donde $S = \{x \in D : g_1(x) \le b_1, \ldots, g_m(x) \le b_m, x_1 \ge 0, \ldots, x_n \ge 0\}$ y $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \ldots, g_m) : D \longrightarrow \mathbb{R}^m$.

Definición 4.1. La función Lagrangiana asociada a (4.1) es

$$L(x, \lambda) = f(x) + \lambda \cdot (b - g(x)), \tag{4.2}$$

donde $\lambda = (\lambda_1, \ldots, \lambda_m)$.

Observar que es este caso la Lagrangiana no incorpora las restricciones de no negatividad. Simplemente asigna multiplicadores al resto de restricciones $g_i(x) \le b_i$.

Teorema 4.2 (Método de Kuhn-Tucker para problemas con restricciones de no negatividad)**.** Supongamos que las funciones $f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$, $g = (g_1, \ldots, g_m) : D \longrightarrow \mathbb{R}^m$ son de clase C^1 en D y que x_0 es un punto regular de[1] (4.1). Si x_0 es una solución del problema 1.1, entonces existe un vector $\lambda_0 = (\lambda_1^0, \ldots, \lambda_m^0) \in \mathbb{R}^m$ tal que

$$\forall i \in \{1, \ldots, n\} \quad \left\{ \begin{array}{rcl} x_i \frac{\partial L}{\partial x_i}(x_0, \lambda_0) & = & 0, \\ \frac{\partial L}{\partial x_i}(x_0, \lambda_0) & \le & 0 \end{array} \right.$$

$$\forall i \in \{1, \ldots, m\} \quad \left\{ \begin{array}{rcl} \lambda_i^0 (b_i - g_i(x_0)) & = & 0, \\ \lambda_i^0 & \ge & 0, \\ g_i(x_0) & \le & b_i. \end{array} \right.$$

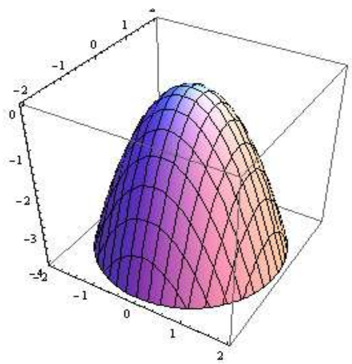
5. Optimización con funciones cóncavas y con funciones convexas

Seat C un subconjunto un conjunto no vacío convexo de $\mathbb{R}^n$. Consideramos de manera independiente los problemas de optimización siguientes:

(1) La función objetivo $f : C \to \mathbb{R}$ es cóncava en C y el problema consiste en

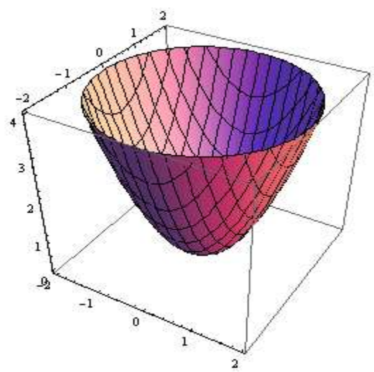
$$\max_{x \in C} f(x)$$

[1] Para ello, basta que la matriz cuyas columnas son los vectores gradiente $\nabla g_1(x_0)^t, \ldots, \nabla g_m(x_0)^t$, tenga rango m; no es necesario considerar las restricciones de no negatividad en esta condición.



(2) La función objetivo $f : C \to \mathbb{R}$ es convexa em C y el problema consiste en

$$\min_{x \in C} f(x)$$



Proposición 5.1. Sea C un subconjunto un conjunto no vacío convexo de $\mathbb{R}^n$. Sea $f : C \to \mathbb{R}$.

(1) si f es cóncava y x_0 es un máximo local de f en C, entonces x_0 es un máximo global de f en C.
(2) si f es convexa y $x_0 \in C$ es un mínimo local de f en C, entonces x_0 es aun mínimo global de f en C.

Proposición 5.2. Sea $C \subset \mathbb{R}^n$ un conjunto no vacío, abierto y convexo, $x_0 \in C$ y sea f de clase C^1 en C.

(1) Si f es cóncava en C, entonces, x_0 es un máximo global de f en C si y sólo si $\nabla f(x_0) = 0$.
(2) Si f es convexa en C, entonces x_0 es un mínimo global de f en C isi y sólo si $\nabla f(x_0) = 0$.

Demostración Si, por ejemplo, f es cóncava, entonces para cada $x \in D$ tenemos que

$$f(x) \leq f(p) + \nabla f(p)(x - p) = f(p)$$

ya que $\nabla f(p) = 0$.

Nota 5.3. Si una función es estrictamente cóncava (resp. convexa) entonces se verifica que

$$f(x) < f(p) + \nabla f(p)(x - p) = f(p)$$

y vemos que si tiene un máximo (resp. mínimo), entonces éste es único. También se puede demostrar directamente a partir de la definición, sin necesidad de utilizar las condiciones de primer orden.

Proposición 5.4. Sea $C \subseteq \mathbb{R}^n$ no vací y convexo let $f : C \to \mathbb{R}$. Entonces

(1) El conjunto formado por los mínimos globales de una función convexa, es convexo.
(2) El conjunto formado por los máximos globales de una función cóncava, es convexo.

Proof Probaremos sólo la primera afirmación, pues la segunda es inmediata considerando $-f$, que es una función cóncava si f es convexa. Sea M el conjunto de minimizadores de f. Si $M = \emptyset$, entones el resultado es trivial.. Sean $x_1, x_2 \in M$y sea $\lambda \in [0,1]$. Tenemos $m = f(x_i) \leq f(x)$, para todol $x \in M$, para $i = 1, 2$. Dado que f es convexa, $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) = m$, por lo que $\lambda x_1 + (1-\lambda)x_2$ es un mínimo global de f, luego $\lambda x_1 + (1-\lambda)x_2 \in M$ y M es convexo.

Teorema 5.5. Sean $C \subseteq \mathbb{R}^n$ un conjunto no vacií y convexo, $f : C \to \mathbb{R}$ y sea x_0 un punto interior de C. Entonces

(1) Si f es convexa y x_0 es un máximo global de f en C, entonces f es constante.
(2) Si f es cóncava y x_0 is mínimo global de f en C, entoncesn f es constante.

Definición 5.6. Sea $C \subseteq \mathbb{R}^n$ un conjunto convexo. Un punto $x_0 \in C$ es un **vértice** de C si $x_0 = \lambda x_1 + (1-\lambda)x_2$ para $x_1, x_2 \in C$ y $\lambda \in [0,1]$, implica $\lambda = 0$ o $\lambda = 1$.

Es decir, un punto del conjunto es un vértice, si no puede expresarse como combinación convexa no trivial de otros dos puntos del conjunto. Notar que los vértices son siempre puntos frontera que pertenecen al conjunto (pero no todo punto frontera es un vértice, ¿por qué?). Por ejemplo, consideremos los conjuntos

$$\begin{aligned}
C_1 &= \{(x,y) \in \mathbb{R}^2 \,:\, x + y < 1\}, \\
C_2 &= \{(x,y) \in \mathbb{R}^2 :\, x + y \leq 1\}, \\
C_3 &= \{(x,y) \in \mathbb{R}^2 :\, x + y \leq 1, x \geq 0, y \geq 0\}, \\
C_4 &= \{(x,y) \in \mathbb{R}^2 :\, x^2 + y^2 \leq 1\}.
\end{aligned}$$

Todos son conjuntos convexos. Ni C_1 ni C_2 tienen vértices. C_3 tiene tres vértices, los puntos $(0,0)$, $(0,1)$ y $(1,0)$. C_4 tiene un número infinito de vértices (la circunferencia $x^2 + y^2 = 1$).

Teorema 5.7. Sea $C \subseteq \mathbb{R}^n$ un conjunto no vacío, convexo y compacto, y sea $f : C \to \mathbb{R}$ continua. Entonces C tiene vértices y además:

(1) Si f es convexa, entonces f alcanza máximo global en alguno de los vértices de C.
(2) Si f es cóncava, entonces f alcanza mínimo global en alguno de los vértices de C.

Ejemplo 5.8. La función $f(x,y) = -x^2 - y^2 + \ln(1+x+y)$ es (estrictamente) cóncava en el conjunto compacto y convexo C_3 descrito anteriormente. Las derivadas de f son

$$\nabla f(x,y) = \left(-2x + \frac{1}{1+x+y}, -2y + \frac{1}{1+x+y}\right)$$

$$Hf(x,y) = \begin{pmatrix} -2 - \dfrac{1}{(1+x+y)^2} & -\dfrac{1}{(1+x+y)^2} \\ -\dfrac{1}{(1+x+y)^2} & -2 - \dfrac{1}{(1+x+y)^2} \end{pmatrix}.$$

Los menores principales de la matriz Hessiana de f son: $\Delta_1 = -2 - \frac{1}{(1+x+y)^2} < 0$, y $\Delta_2 = 4 + \frac{4}{(1+x+y)^2} > 0$.

Los vértices de C_3 fueron determinados como $(0,0)$, $(0,1)$ y $(1,0)$. Por el teorema anterior, los mínimos globales de f en C_3 se encuentran entre estos puntos. Dado que $f(0,0) = 0$, $f(0,1) = f(1,0) = -1 + \ln 2 < -1 + \ln e = -1 + 1 = 0$, los puntos $(0,1)$ y $(1,0)$ son mínimos globales de f en C_3.

Notar que el teorema no dice nada sobre los máximos de f. Podrían localizarse sobre el interior o sobre la frontera de C_3. Para hallarlos, habría que resolver el problema de KT correspondiente. Sin embargo, en este caso, las condiciones de primer orden de optimalidad

$$\frac{\partial f}{\partial x}(x,y) = -2x + \frac{1}{1+x+y} = 0,$$

$$\frac{\partial f}{\partial y}(x,y) = -2y + \frac{1}{1+x+y} = 0,$$

tienen como única solución el punto $(\frac{1}{4}(\sqrt{5}-1), \frac{1}{4}(\sqrt{5}-1))$, que pertenece al conjunto C_3, pues sus componentes son no negativas y suman $\frac{1}{2}(\sqrt{5}-1) \leq 1$. Por tanto, dicho punto es el único máximo global de f en C_3 (y en todo $\mathbb{R}^2_+$).