1 Optimización restringida

1.1 Una receta de cocina para resolver problemas de optimización restringida

Esta receta tiene 7 pasos.

Paso 0 : Escribir el problema en la forma canónica.

Esto es convertir el problema en uno que se vea así:

$$\begin{aligned} \max_{x \in \mathbb{R}^k} \quad & f(x) \\ s.a \quad & \\ & h_i(x) = c_i \qquad i = 1, \ldots, n \\ & g_j(x) \leq c_j \qquad j = 1, \ldots, m \end{aligned} \tag{1}$$

Paso 1 : Formar el Lagrangeano agregando λ_i $i = 1, \ldots, n$ y λ_j $j = 1, \ldots, m$ para tener una función de $k + n + m$ variables:

$$L = f(x) - \sum_{i=1}^{n} \lambda_i (h_i(x) - c_i) - \sum_{j=1}^{m} \lambda_j (g_j(x) - c_j)$$

Paso 2 : Expresar las condiciones de primer orden para cada x_h $h = 1, \ldots, k$.

$$\frac{\partial L}{\partial x_h} = \frac{\partial f}{\partial x_h} - \sum_{i=1}^{n} \lambda_i \frac{\partial h_i(x)}{\partial x_h} - \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(x)}{\partial x_h} = 0$$

Paso 3 : Agregar las $n + m$ restricciones.

$$\begin{aligned} h_i(x) &= c_i \qquad i = 1, \ldots, n \\ g_j(x) &\leq c_j \qquad j = 1, \ldots, m \end{aligned}$$

Paso 4: Para los m λ_j, agregar las restricciones de no negatividad.

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

Paso 5: Agregar las condiciones de holgura complementaria.

$$
\begin{aligned}
\lambda_j(g_j - c_j) &= 0 \qquad i = 1, \ldots, m \\
\lambda_i(h_i - c_i) &= 0 \qquad i = 1, \ldots, n \text{ (Estas no aportan mucho)}
\end{aligned}
$$

Paso 6: Mezcle todos los ingredientes y resuelva por ensayo y error. Si encuentra una solución, es una posible solución del problema.

1.2 ¿Por qué funciona la receta?

Teorema 1 (Lagrange) *En el problema:*

$$
\begin{aligned}
\max_{x \in \mathbb{R}^n} \quad & f(x) \\
s.a \quad & \\
& g_i(x) = c_i \qquad i = 1, \ldots, m
\end{aligned}
$$

Si:

1. *f y g_i $i = 1, \ldots, n$ son diferenciables.*
2. *$\bar{x} \in C$ es una solución.*
3. *La matriz de $n \times m$*

$$
M = \begin{bmatrix} \frac{\partial g_1(\bar{x})}{\partial x_1} & \cdots & \frac{\partial g_1(\bar{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(\bar{x})}{\partial x_1} & \cdots & \frac{\partial g_m(\bar{x})}{\partial x_n} \end{bmatrix}
$$

tiene rango m (condición de calificación de restricciones).

Entonces, existen m escalares $\lambda_i \in \mathbb{R}$, llamados multiplicadores de Lagrange, tales que:

$$
\frac{\partial f(\bar{x})}{\partial x_j} = \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(\bar{x})}{\partial x_j} \qquad \forall j = 1, \ldots, n
$$

Dem.

La condición de calificación de restricciones, garantiza que $\bar{x}$ también es solución del siguiente problema (que no es sino una linearización, en torno a $\bar{x}$ del problema original)

$$\begin{array}{rl} \max\limits_{x \in \mathbb{R}^n} & f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) \\ s.a & \\ & \nabla g_i(x) \cdot (x - \bar{x}) = c_i \qquad i = 1, \ldots, m \end{array}$$

Intuitivamente, si $\bar{x}$ es una solución del problema, entonces cualquier dirección de desplazamiento $z \in \mathbb{R}^n$ tal que no tiene efectos de primer orden en las restricciones, tampoco puede tenerlos en la función objetivo. Es decir si $\bar{x}$ es solución, no debe ser posible aumentar el valor de la función objetivo sin violar alguna restricción.

Por lo tanto, cualquier dirección z tal que:

$$\nabla g_i(\bar{x}) \cdot z = 0 \qquad \forall i = 1, \ldots, m$$

implica que:

$$\nabla f(\bar{x}) \cdot z = 0$$

Por lo tanto la matriz de $(m+1) \times n$:

$$E = \begin{bmatrix} \frac{\partial f(\bar{x})}{\partial x_1} & \cdots & \frac{\partial f(\bar{x})}{\partial x_n} \\ \frac{\partial g_1(\bar{x})}{\partial x_1} & \cdots & \frac{\partial g_1(\bar{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(\bar{x})}{\partial x_1} & \cdots & \frac{\partial g_m(\bar{x})}{\partial x_n} \end{bmatrix}$$

satisface que:

$$z / E \cdot z = 0 \Leftrightarrow \nabla g_j \cdot z = 0$$

luego:

$$M \cdot z = 0$$

Como esto sólo es posible si E y M tienen el mismo rango, entonces agregar la fila $\nabla f(\bar{x})'$ no tuvo efectos en el rango de la matriz. En conclusion $\nabla f(\bar{x})$ es una combinación lineal de los $\nabla g_i(\bar{x})$ y, por lo tanto, existen m escalares tales que:

$$\frac{\partial f(\bar{x})}{\partial x_j} = \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(\bar{x})}{\partial x_j} \qquad \forall j = 1, \ldots, n$$

$\square$

Nótese que esta es una condición necesaria pero no suficiente, es decir, que existan los multiplicadores de Lagrange no implica que el punto sea solución del problema, sino, solamente, que es una posible solución del mismo.

Definición 1 (Hiperplano) *Sean $z \in \mathbb{R}^n$ y $\beta \in \mathbb{R}$, se define el hiperplano $H_{z,\beta}$ como el conjunto de los $x \in \mathbb{R}^n$ tales que:*

$$z{\cdot}x = \beta$$

Este hiperplano divide al espacio en dos mitades, la mitad superior $H^s_{z,\beta}$, compuesta por todos los puntos x_s tales que

$$z{\cdot}x_s > \beta$$

y la mitad inferior $H^i_{z,\beta}$, compuesta por todos los puntos x_i tales que:

$$z{\cdot}x_i < \beta$$

Teorema 2 (Hiperplano separador) *Sea el conjunto convexo y cerrado $B \subset \mathbb{R}^n$ y un punto $x \in \mathbb{R}^n \notin B$, entonces, existe un hiperplano $H_{z,beta}$ tal que, si $B \subset H^s_{z,\beta}$ entonces $x \in H^i_{z,beta}$ y viceversa.*

Dem.

Sea y el punto de B que es más cercano, en el sentido de la norma euclidiana, a x. Definamos $z = x - y$ y $\beta' = z{\cdot}y$, entonces, es fácil ver que $z{\cdot}x > \beta'$. En efecto:

$$z{\cdot}x - \beta = (x-y){\cdot}x - (x-y){\cdot}y = (x-y){\cdot}(x-y) = \|(x-y)\|^2 > 0$$

Además para cualquier $w \in B$ el ángulo entre $(w - y)$ y $(x - y)$ no puede ser agudo (si lo fuera, entonces y no sería el punto de B más cercano a x). Luego,

$$(x-y){\cdot}(w-y) < 0$$

es decir:

$$(x-y){\cdot}w - (x-y){\cdot}y < 0$$

o, lo que es lo mismo,

$$z{\cdot}w < \beta'$$

Sea ahora $\beta = \beta' + \varepsilon$ con $\varepsilon > 0$ suficientemente pequeño para que $z{\cdot}x > \beta$. Entonces, el hiperplano $H_{z,\beta}$ es tal que $B \in H^i_{z,\beta}$ y $x \in H^s_{z,\beta}$. □

Teorema 3 (Kuhn–Tucker) *Supongamos que $\bar{x} \in C$ es una solución del problema:*

$$\begin{array}{ll} \max\limits_{x \in \mathbb{R}^n} & f(x) \\ s.a & \\ & g_i(x) \le c_i \qquad i = 1, \ldots, m \end{array}$$

si se satisface la condición de calificación de restricciones, entonces existen m multiplicadores de Lagrange $\lambda_i \in \mathbb{R}^+$ tales que:

1. *Para todo $j = 1, \ldots, n$:*

$$\frac{\partial f(\bar{x})}{\partial x_j} = \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(\bar{x})}{\partial x_j}$$

2. *Para todo $i = 1, \ldots, m$:*

$$\lambda_i (g_i(\bar{x}) - c_i) = 0$$

Dem.

Intuitivamente, demostraremos que, para toda dirección de desplazamiento $z \in \mathbb{R}^n$ tal que $\nabla g_i(\bar{x}) \cdot z \leq 0$ $i = 1, \ldots, m$ (es decir, que satisface las restricciones) entonces $\nabla f(\bar{x}) \cdot z \leq 0$, es decir, no es una dirección de crecimiento de $f(x)$. Es decir, si hay una dirección en la cual la función objetivo puede aumentar, entonces en dicha dirección debe violarse, al menos, una restricción.

El teorema establece que, para que $\bar{x}$ sea una solución del problema, $\nabla f(\bar{x})$ debe estar en el cono definido por las restricciones activas:

$$\Gamma = \left\{ y \in \mathbb{R}^n / y = \sum_j \lambda_j \nabla g_j(\bar{x}) \qquad \forall \lambda_j \geq 0 \right\}$$

Supongamos que esto no es así, es decir, $\nabla f(x) \notin \Gamma$, entonces, por el teorema del hiperplano separador, existen un vector $z \in \mathbb{R}^n$ y un número $\beta \in \mathbb{R}$ tal que $\nabla f(x) \cdot z \geq \beta$ y $y \cdot z < \beta$ $\forall y \in \Gamma$. Como $0 \in \Gamma$, esto implica $\beta > 0$, por lo tanto, $\nabla f(\bar{x}) \cdot z \geq 0$.

Además, $\forall y \in \Gamma, \theta \in \mathbb{R}^+$, $\theta y \in \Gamma$, por la definición de cono. De donde se obtiene que:

$$\theta y \cdot z < \beta \qquad \forall \theta \in \mathbb{R}^+$$

En particular, sea $\theta = \left(1 + \frac{\beta}{y \cdot z}\right)$ entonces:

$$\left(1 + \frac{\beta}{y \cdot z}\right) y \cdot z = y \cdot z + \beta < \beta$$

es decir, $y \cdot z < 0$ y, por lo tanto, $\nabla g_i(\bar{x}) \cdot z \leq 0$ para todo i y $\nabla f(\bar{x}) \cdot z \geq 0$ (es decir, existe una dirección de crecimiento factible que permite incrementar la función) lo que contradice a la linearización del problema que se desprende de la condición de calificación de restricciones. $\square$