



1

Repaso de optimización

El problema general de optimización se refiere a la búsqueda de los valores más altos o más bajos que un función f alcanza en un determinado conjunto.

Es útil recordar que una función real $f : X \rightarrow \mathbb{R}$ es una relación que asocia a un valor $x \in X$ un único número real $y \in \mathbb{R}$. Se escribe $y = f(x)$. X es el dominio de la función e $Y \subset \mathbb{R}$ su recorrido.

El máximo¹ de una función es el elemento $f(x^*) \in \mathbb{R}$ que satisface lo siguiente:

$$f(x^*) > f(x) \quad \forall x \in X, \text{ con } x \neq x^*$$

En este caso, $x^* \in \mathbb{R}$ es el argumento de la maximización de f , denotado $x^* = \arg \max f(x)$.

El mínimo de una función, por su parte, es el elemento $f(x^*) \in \mathbb{R}$ que satisface:

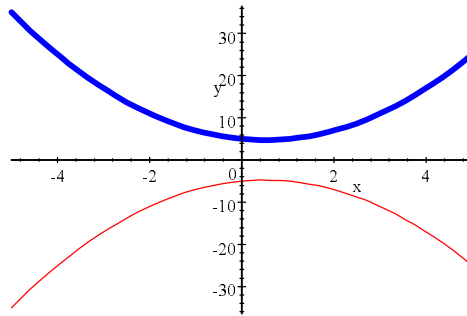
$$f(x^*) < f(x) \quad \forall x \in X, \text{ con } x \neq x^*$$

En este caso, $x^* \in \mathbb{R}$ es el argumento de la minimización de f , denotado $x^* = \arg \min f(x)$.

Es importante distinguir entre la función y su argumento. En un problema de optimización, la función recibe el nombre de objetivo.

Observe que $-f(x)$ es el “reflejo” de $f(x)$, donde la línea del agua está en el 0. Por ejemplo, a continuación se ilustran el gráfico de $f(x) = x^2 - x + 5$ (en azul) y su “reflejo” $-x^2 + x - 5$

¹En general, es posible que exista más de un máximo o mínimo, por lo que correspondería hablar de “un” y no “el”, y la condición se satisfaría con desigualdad débil y no estricta. Sin embargo, para efectos de esta exposición relajaremos el supuesto de unicidad sólo al final del capítulo.



Así, es inmediato verificar que si $x^* = \arg \min f(x)$, entonces $x^* = \arg \max \{-f(x)\}$. En virtud de ello, nos podemos concentrar exclusivamente en la maximización.

En lo que sigue, supondremos que f es una función diferenciable.

1.1 Maximización sin restricciones

Empezamos analizando un problema del tipo:

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

Como mencionamos anteriormente, si $f(x^*)$ es el máximo, no puede haber otro $f(x)$ de mayor valor en ninguna parte. En particular, localmente, es decir, en la vecindad de x^* . Esto permite usar la derivada de la función para facilitar la búsqueda de máximos. El método de búsqueda de un máximo utilizando el cálculo explota esta observación de la siguiente forma: si x^* es un máximo, entonces cualquier movimiento infinitesimal en cualquier dirección debiera traer como consecuencia un menor valor de y .

En el caso más sencillo de una función de una variable, sabemos que y aumenta o disminuye en respuesta a una alteración marginal en x de acuerdo a:

$$dy = \frac{\partial f}{\partial x} dx = f'(x) dx$$

Luego, resulta inmediato que un máximo debe satisfacer:

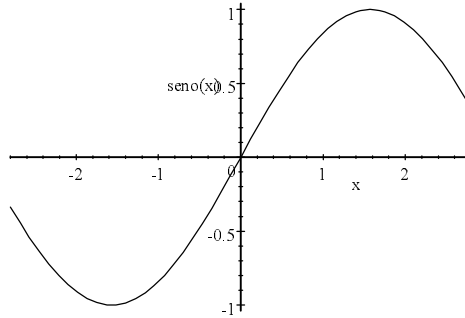
$$\frac{\partial f}{\partial x} = f'(x) = 0 \tag{1.1}$$

Si esto no fuese así, siempre podríamos encontrar al menos una dirección en la cual conseguiríamos aumentar y . En efecto,

$$dy > 0 \Rightarrow \frac{\partial f}{\partial x} dx > 0$$

de manera que si $\frac{\partial f}{\partial x} > 0$, con un movimiento a la derecha de x^* ($dx > 0$) se consigue una mejora en el objetivo, y si $\frac{\partial f}{\partial x} < 0$, basta con moverse infinitesimalmente a la izquierda de x^* ($dx < 0$) para mejorar.

Pero por supuesto esto no es suficiente, porque la misma condición puede ser usada para evitar una caída de y ($dy < 0$). Por ejemplo, la primera derivada de la función $y = \text{sen}(x)$ tiene una primera derivada igual a 0 en $\frac{\pi}{2}$ y en $-\frac{\pi}{2}$, y obviamente -1 no es el máximo, como se aprecia en el gráfico:



Lo que falta es verificar es que al moverse en cualquier dirección, el valor de y caiga. Aquí aparece la segunda observación crucial: la variación de y debe ocurrir a tasas decrecientes, es decir:

$$d^2y = \frac{\partial^2 f}{\partial x^2} (dx)^2 = f''(x) (dx)^2 < 0 \quad (1.2)$$

¿Por qué? Observe el gráfico de izquierda a derecha ($dx > 0$). Las diferenciales son negativas y decrecientes, luego positivas y crecientes. Esto tiene que ser así alrededor de un mínimo: a su izquierda deben haber valores mayores, luego la diferencial en cualquiera de esos puntos debe ser negativa; y a tasas decrecientes para hacer una transición “suave” hacia la diferencial positiva (de lo contrario, la función no sería diferenciable en el punto. Esta salvedad debiera dejar claro que este método no es útil en toda circunstancia). Simétricamente, al aproximarse a un máximo se tienen diferenciales positivas y decrecientes y luego negativas y crecientes. Así, podemos diferenciar un mínimo de un máximo por la segunda diferencial en el punto.

Entonces, (1.1) es la condición de primer orden y (1.2) de segundo orden. La condición de primer orden es necesaria, pero no suficiente (recuerde que un mínimo también la satisface) para obtener un **máximo local interior**. Es suficiente para un máximo local interior que ambas se satisfagan simultáneamente.

Enfatizamos la palabra local porque la búsqueda se restringió a la vecindad del punto. Es posible que otros puntos satisfagan ambas condiciones; el máximo global en ese caso se obtiene por comparación directa de los valores de $f(x)$ entre los candidatos.

Enfatizamos también la palabra interior, porque es posible que el máximo en un dominio acotado ocurra en los extremos. Por ejemplo, el máximo de $3x^2 + 1$ en el intervalo $[0, 1]$ ocurre en el punto $x = 1$. En este punto no se cumple ni la condición de primer orden ni de segundo, por lo que no pueden ser ni necesarias ni suficientes en general.

Con más de una variable, la intuición se mantiene. La única diferencia es que no basta con chequear una dimensión, sino que se hace necesario verificar movimientos en toda dirección posible. Así,

$$\begin{aligned} y &= f(x_1, x_2, \dots, x_n) \\ dy &= \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \dots + \frac{\partial f}{\partial x_n} dx_n = 0 \\ &= \sum_{i=1}^n f_i dx_i = 0 \end{aligned} \tag{1.3}$$

$$\begin{aligned} d^2y &= \frac{\partial^2 f}{\partial x_1^2} (dx_1)^2 + \frac{\partial^2 f}{\partial x_1 \partial x_2} dx_1 dx_2 + \dots + \frac{\partial^2 f}{\partial x_1 \partial x_n} dx_1 dx_n \\ &\quad + \frac{\partial f}{\partial x_2 \partial x_1} dx_2 dx_1 + \frac{\partial^2 f}{\partial x_2^2} (dx_2)^2 + \dots + \frac{\partial^2 f}{\partial x_n \partial x_2} dx_2 dx_n + \dots \\ &\quad + \frac{\partial f}{\partial x_n \partial x_1} dx_n dx_1 + \frac{\partial^2 f}{\partial x_n \partial x_2} dx_n dx_2 + \dots + \frac{\partial^2 f}{\partial x_n^2} (dx_n)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n f_{ij} dx_i dx_j < 0 \end{aligned} \tag{1.4}$$

Como antes, un punto $\mathbf{x}^* \in \mathbb{R}^n$ (x en negrita denota un vector, y sin negrita un escalar) es un máximo de $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ si desviaciones infinitesimales no cambian el valor del objetivo: $dy = 0$. En la ecuación (1.3) vemos que esto se cumple al moverse en cualquier dirección si todas las primeras derivadas parciales de la función son 0, es decir, la condición de primer orden es:

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 0 \\ \frac{\partial f}{\partial x_2} &= 0 \\ &\vdots \\ \frac{\partial f}{\partial x_n} &= 0 \end{aligned} \tag{1.5}$$

Por su parte, la condición de segundo orden es ligeramente más compleja de verificar. Consideremos primero lo que ocurre al mover una variable a la vez. Como las direcciones aparecen en forma cuadrática (dx^2), es claro que la única forma de

cumplir la condición es que las segundas derivadas propias sean negativas ($f_{ii} < 0$), como antes. Pero al considerar movimientos de dos o más variables a la vez, nos empezamos a encontrar con nuevas condiciones. Así, por ejemplo, en el caso de dos variables,

$$d^2y = f_{11}(dx_1)^2 + 2f_{12}dx_1dx_2 + f_{22}(dx_2)^2 < 0$$

Esto debe cumplirse para cualquier dirección que escojamos, por ejemplo las siguientes:

$$\begin{aligned} dx_1 = 0 &\Rightarrow d^2y = f_{22}(dx_2)^2 < 0 \Rightarrow f_{22} < 0 \\ dx_2 = 0 &\Rightarrow d^2y = f_{11}(dx_1)^2 < 0 \Rightarrow f_{11} < 0 \\ dx_1 = \sqrt{-f_{22}}, dx_2 = \sqrt{-f_{11}} &\Rightarrow d^2y = -f_{11}f_{22} + 2f_{12}\sqrt{f_{11}f_{22}} - f_{22}f_{11} < 0 \\ &\Rightarrow f_{12} < \sqrt{f_{11}f_{22}} \end{aligned}$$

La forma arbitraria en que escogimos esta última dirección puede hacer dudar de si no hay más implicancias del requisito $d^2y < 0$; la verdad es que no. En efecto, consideremos primero las direcciones en que $dx_1dx_2 \geq 0$:

$$\begin{aligned} f_{12} &< \sqrt{f_{11}f_{22}} \Rightarrow \\ f_{11}(dx_1)^2 + 2f_{12}dx_1dx_2 + f_{22}(dx_2)^2 &< f_{11}(dx_1)^2 + 2\sqrt{f_{11}f_{22}}dx_1dx_2 + f_{22}(dx_2)^2 \\ &= -\left(\sqrt{-f_{11}}dx_1 - \sqrt{-f_{22}}dx_2\right)^2 \\ &< 0 \end{aligned}$$

Por otra parte, si $dx_1dx_2 < 0$, tenemos:

$$\begin{aligned} f_{12} &< \sqrt{f_{11}f_{22}} \Rightarrow \\ f_{11}(dx_1)^2 - 2f_{12}dx_1dx_2 + f_{22}(dx_2)^2 &< f_{11}(dx_1)^2 - 2\sqrt{f_{11}f_{22}}dx_1dx_2 + f_{22}(dx_2)^2 \\ &= -\left(\sqrt{-f_{11}}dx_1 + \sqrt{-f_{22}}dx_2\right)^2 \\ &< 0 \end{aligned}$$

Una manera compacta de escribir la condición anterior es que la matriz de segundas derivadas (también conocida como el Hessiano de f) sea negativa definida:

$$\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} \text{ neg. def.}$$

Recuerde que una matriz H es negativa definida si los determinantes de los menores alternan signo, empezando en negativo. Recuerde también que los menores son las matrices que se forman eliminando filas y columnas de la matriz principal.

Partiendo del extremo superior izquierdo, el primero es la primera entrada. El segundo extremo se forma agregando al primero la fila y la columna contiguas. El tercero de la misma forma, a partir del segundo, y así sucesivamente.

Por ejemplo, en el caso de dos variables, H negativa definida se traduce en:

$$\begin{aligned} |H_1| &= |f_{11}| < 0 \Leftrightarrow f_{11} < 0 \\ |H_2| &= \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} > 0 \Leftrightarrow f_{11}f_{22} - (f_{12})^2 > 0 \end{aligned}$$

Observe que la condición $f_{22} < 0$ también se deduce de la condición anterior:

$$\begin{aligned} f_{11}f_{22} > (f_{12})^2 \wedge f_{11} < 0 \\ \Rightarrow f_{22} < \frac{(f_{12})^2}{f_{11}} < 0 \end{aligned}$$

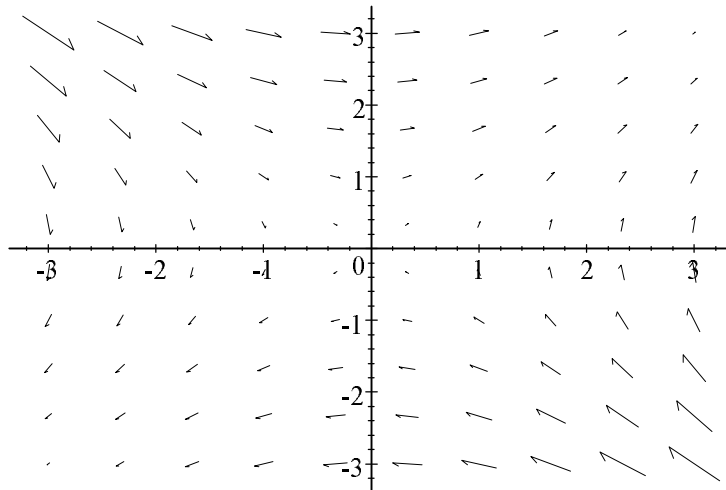
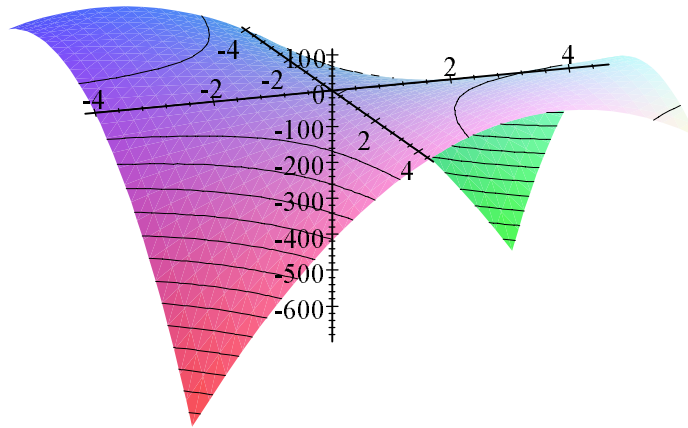
De manera que la expresión “ H negativa definida” es una forma compacta de decir “ $f_{11}, f_{22} < 0$ y $f_{12} < \sqrt{f_{11}f_{22}}$ ” en el caso de dos variables. En el caso general en que hay n variables, la condición de segundo orden es “ H negativa definida”, expresión que sintetiza una serie de requisitos sobre las derivadas cruzadas de f .

Observe que $dy = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}' \begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix}$ y $d^2y = \begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix}' \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix}$. Como el problema de optimización se trata de “jugar” con los movimientos $\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix}$ de modo de obtener una condición sobre y , es natural que la solución se exprese en términos de la gradiente $\begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$ y el Hessiano $\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$. En general, escribamos $dy = G'd\mathbf{x}$ y $d^2y = d\mathbf{x}'Hd\mathbf{x}$.

Resumen 1 Para un máximo interior local, las condiciones necesarias y suficientes son:

1. Gradiente igual a 0 ($G = 0$)
2. Hessiano negativo definido (H neg. def.)

Ejemplo 1 La función $20x_1x_2 - x_1^2x_2^2$ satisface la condición de primer orden en $\{x_1 = 0, x_2 = 0\}$ y en $\left\{x_1 = \frac{10}{x_2}, x_2 = x_2\right\}$. El primer gráfico corresponde a f en sus tres dimensiones. El segundo corresponde a la gradiente de f :



1.2 Maximización con restricciones

La maximización con restricciones se refiere al mismo problema anterior, con la salvedad de que la búsqueda se restringe a un subconjunto propio del dominio original de la función. Para facilitar la exposición, normalmente se distinguen dos clases de restricciones: de igualdad y de desigualdad. La restricción de desigualdad es la más general, y corresponde a acotar arbitrariamente el dominio de la función objetivo. La de igualdad es aquella en la que el conjunto de puntos en los que se permite buscar

pueden expresarse por medio de una función del tipo $g(x_1, x_2, \dots, x_n) = b$. Siguiendo la práctica común, comenzaremos por esta última.

1.2.1 Restricciones de igualdad

Para abordar este problema hay en general dos estrategias posibles; la elección se hace sencillamente por conveniencia.

La primera estrategia reduce la dimensión del problema. En efecto, el problema inicial

$$\begin{aligned} \max_{\mathbf{x} \in \mathbb{R}^n} f(x_1, x_2, \dots, x_n) \text{ sujeto a} \\ b = g(x_1, x_2, \dots, x_n) \end{aligned} \quad (1.6)$$

se transforma obteniendo de $b = g(x_1, x_2, \dots, x_n)$ una expresión para alguna variable, digamos $x_2 = h(x_1, x_3, \dots, x_n)$, y reemplazándola en la función objetivo para obtener:

$$\max_{\mathbf{x} \in \mathbb{R}^{n-1}} f(x_1, h(x_1, x_3, \dots, x_n), \dots, x_n) \quad (1.7)$$

El nuevo problema se trata como lo explica la sección anterior. Este método es sencillo, pero a veces puede resultar impracticable por la imposibilidad de despejar una variable de la restricción, o simplemente engorroso. En ocasiones, el segundo método es preferido porque entrega información adicional sobre las características del óptimo que es útil en determinadas aplicaciones.

La segunda estrategia, conocida como el **método de Lagrange**, de hecho aumenta la dimensión del problema al transformarlo en:

$$\max_{\mathbf{x} \in \mathbb{R}^n, \lambda \in \mathbb{R}} f(x_1, x_2, \dots, x_n) + \lambda [b - g(x_1, x_2, \dots, x_n)] \quad (1.8)$$

donde el escalar λ es considerado como una variable más. Observe lo siguiente:

1. Si la restricción es de hecho satisfecha, la nueva función $\mathcal{L} = f(x_1, x_2, \dots, x_n) + \lambda [b - g(x_1, x_2, \dots, x_n)]$ alcanza el mismo máximo que el objetivo inicial.
2. Al considerar a λ como una variable de elección, la condición de primer orden va a exigir la satisfacción de la restricción: $\frac{\partial \mathcal{L}}{\partial \lambda} = b - g(x_1, x_2, \dots, x_n) = 0$, de manera que cualquier solución que la maximización de la nueva función (el lagrangeano) va a pertenecer al conjunto de puntos admisible.

La condición de primer orden es la misma (y por las mismas razones) que en la sección anterior, vale decir, la gradiente de \mathcal{L} debe ser cero pues de lo contrario podríamos encontrar formas de aumentar f .

La condición de segundo orden, en cambio, es diferente. La razón es que al restringir la búsqueda a los puntos que satisfagan $g(x_1, x_2, \dots, x_n) = b$, de hecho eliminamos las restricciones provenientes de pedirlo en direcciones inadmisibles.

Consideremos primero el caso de dos variables. Al añadir la restricción $g(x_1, x_2) = b$, reducimos la dimensión del problema en uno. La restricción implica que no es necesario que movimientos en cualquier dirección produzcan una caída en y , ya que no se nos permite movernos en cualquier dirección. En particular, las direcciones admisibles satisfacen:

$$dg = g_1 dx_1 + g_2 dx_2 = 0 \Rightarrow dx_2 = -\frac{g_1}{g_2} dx_1$$

Entonces, la condición de segundo orden es:

$$\begin{aligned} d^2 y &= f_{11} (dx_1)^2 + 2f_{12} dx_1 dx_2 + f_{22} (dx_2)^2 < 0 \\ \Rightarrow f_{11} (dx_1)^2 + 2f_{12} dx_1 \left(-\frac{g_1}{g_2} dx_1\right) + f_{22} \left(-\frac{g_1}{g_2} dx_1\right)^2 < 0 \\ \Leftrightarrow \left(\frac{dx_1}{g_2}\right)^2 (f_{11} g_2^2 - 2f_{12} g_1 g_2 + f_{22} g_1^2) < 0 \\ \Leftrightarrow (f_{11} g_2^2 - 2f_{12} g_1 g_2 + f_{22} g_1^2) < 0 \end{aligned}$$

lo que corresponde a pedir exclusivamente $|\overline{H}^f| > 0$, donde

$$\overline{H}^f \equiv \begin{pmatrix} f_{11} & f_{12} & g_1 \\ f_{21} & f_{22} & g_2 \\ g_1 & g_2 & 0 \end{pmatrix}$$

es el hessiano orlado (o con bordes) de f , puesto que se construye agregándole un “borde” al hessiano original de f .

Así, en el caso de dos variables hay una sola condición de segundo orden puesto que la búsqueda se reduce a una línea, tal como en el caso de optimización sin restricciones en una variable. Esto no podría ser de otra forma, toda vez que la primera estrategia de solución que discutimos de hecho consiste en reducir el problema a esa categoría.

Ahora bien, con más de dos variables (o en general, si m es el número de restricciones, con $n - m \geq 2$), el problema obviamente se complica porque ya no se busca en una línea sino en conjuntos más complicados y surgen restricciones adicionales. En general, entonces, tenemos:

Resumen 2 *El problema*

$$\begin{aligned} & \max_{\mathbf{x} \in \mathbb{R}^n} f(x_1, x_2, \dots, x_n) \text{ sujeto a} \\ b^1 &= g^1(x_1, x_2, \dots, x_n) \\ & \vdots \\ b^m &= g^m(x_1, x_2, \dots, x_n) \end{aligned}$$

ó

$$\max_{\mathbf{x} \in \mathbb{R}^n, \lambda \in \mathbb{R}^m} \mathcal{L} = f(x_1, x_2, \dots, x_n) + \sum_{j=1}^m \lambda_j [b^j - g^j]$$

tiene como solución:

1. (CPO) gradiente de \mathcal{L} igual a 0.
2. (CSO) La secuencia de $(n - m)$ menores orlados $|\overline{H}_{m+1}|, |\overline{H}_{m+2}|, \dots, |\overline{H}_n|$ alternan signo, empezando con $(-1)^{1+m}$.

Es importante notar que el multiplicador de Lagrange tiene la interpretación del aporte de una unidad del recurso restringido al objetivo. En efecto,

$$\begin{aligned} \frac{d\mathcal{L}(\mathbf{x}^*)}{db} &= \sum_{i=1}^n \left. \frac{\partial f(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}^*} \frac{\partial x_i^*}{\partial b} + \lambda + \lambda \left[\sum_{i=1}^n \left. \frac{\partial g(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}^*} \frac{\partial x_i^*}{\partial b} \right] \\ &= \lambda + \sum_{i=1}^n \frac{\partial x_i^*}{\partial b} \left[\left. \frac{\partial f(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}^*} + \lambda \sum_{i=1}^n \left. \frac{\partial g(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}^*} \right] \\ &= \lambda \end{aligned}$$

donde el último paso surge de observar que en \mathbf{x}^* la condición de primer orden se satisface. De manera que el valor del multiplicador nos entrega información sobre qué tan valioso es el recurso limitante.

1.2.2 Restricciones de desigualdad

Finalmente, analizamos el problema de la forma:

$$\begin{aligned} &\max_{\mathbf{x} \in \mathbb{R}} f(x_1, x_2, \dots, x_n) \text{ sujeto a} \\ b^1 &\geq g^1(x_1, x_2, \dots, x_n) \\ &\vdots \\ b^m &\geq g^m(x_1, x_2, \dots, x_n) \end{aligned}$$

en que el conjunto de restricciones nuevamente reduce el dominio de la función, pero no limitadas a funciones sino que ahora permitiendo la delimitación de áreas (o volúmenes, o lo que corresponda de acuerdo a la dimensión del problema).

Consideremos primero el caso de una restricción. En general, dos cosas pueden suceder: o la restricción se cumple con igualdad, o lo hace con desigualdad estricta. Si el óptimo irrestricto se encuentra dentro del área encerrada por la restricción, entonces la restricción se satisface “con holgura”, y el hecho de que exista no altera en absoluto

el problema. Por ejemplo, en el problema de la caja, obviamente $x \leq 5$ para que definiera un volumen; pero no fue necesario incorporar explícitamente la restricción porque ésta se satisfizo automáticamente. Ahora bien, si el óptimo irrestricto se encuentra fuera de lo permitido por las restricciones, entonces lo natural es que la restricción se satisfaga con igualdad.

Lo anterior está estrechamente relacionado con el valor del multiplicador de Lagrange: si la restricción se satisface con holgura, entonces un pequeño aumento en la restricción no afecta en absoluto el máximo valor alcanzable (pues de hecho ya sobra), de manera que el multiplicador es 0. Si no se satisface con holgura, ese hecho debiera reflejarse en el valor de λ .

Esta idea se puede expresar complementando el método de Lagrange. El problema

$$\max_{\mathbf{x} \in \mathbb{R}^n, \lambda \in \mathbb{R}^m} \mathcal{L} = f(x_1, x_2, \dots, x_n) + \sum_{j=1}^m \lambda_j [b^j - g^j(x_1, x_2, \dots, x_n)]$$

tiene como condición de primer orden lo siguiente:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_i} &= f_i - \sum_{j=1}^m \lambda_j g_i^j = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_j} &= b^j - g^j(x_1, x_2, \dots, x_n) \geq 0 \quad \text{con holgura complementaria } \lambda_j \frac{\partial \mathcal{L}}{\partial \lambda_j} = 0 \end{aligned}$$

La condición de holgura complementaria resume lo señalado anteriormente: o $\lambda_j = 0$, es decir, la restricción no es operativa, en cuyo caso es perfectamente posible que $b^j - g^j(x_1, x_2, \dots, x_n) < 0$, o bien $\lambda_j > 0$, vale decir, la restricción afecta el máximo valor alcanzable del objetivo y, por tanto, debe satisfacerse con igualdad. Observe que si todas las restricciones se satisfacen con holgura, obtenemos la misma condición de gradiente nula que en un problema sin restricciones.

Respecto de las condiciones de segundo orden, baste decir que dependen de si las restricciones se satisfacen con o sin holgura y, por tanto, se prosigue como se describe en las secciones anteriores.

Una restricción que es muy frecuente en aplicaciones en economía es la no negatividad de las variables de elección:

$$\begin{aligned} &\max_{\mathbf{x} \in \mathbb{R}^n} f(x_1, x_2, \dots, x_n) \text{ sujeto a} \\ &0 \leq x_1, x_2, \dots, x_n \end{aligned}$$

Éste es un caso particular del anterior, pero su forma simple permite una solución que prescinde de los multiplicadores, usando la siguiente condición de primer

orden:

$$\frac{\partial f}{\partial x_i} \leq 0 \quad \text{con holgura complementaria } x_i \frac{\partial f}{\partial x_i} = 0$$

Si ambas clases de restricciones se dan simultáneamente, tenemos que:

$$\begin{aligned} & \max_{\mathbf{x} \in \mathbb{R}^n} f(x_1, x_2, \dots, x_n) \text{ sujeto a} \\ b^1 & \leq g^1(x_1, x_2, \dots, x_n) \\ & \vdots \\ b^m & \leq g^m(x_1, x_2, \dots, x_n) \\ 0 & \leq x_1, x_2, \dots, x_n \end{aligned}$$

formamos el lagrangeano:

$$\max_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L} = f(x_1, x_2, \dots, x_n) + \sum_{j=1}^m \lambda_j [b^j - g^j(x_1, x_2, \dots, x_n)]$$

que tiene como condición de primer orden:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_i} &= f_i - \sum_{j=1}^m \lambda_j g_i^j \leq 0 \quad \text{con holgura complementaria } x_i \frac{\partial \mathcal{L}}{\partial x_i} = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_j} &= b^j - g^j(x_1, x_2, \dots, x_n) \geq 0 \quad \text{con holgura complementaria } \lambda_j \frac{\partial \mathcal{L}}{\partial \lambda_j} = 0 \end{aligned}$$

Este conjunto de condiciones, conocido como las **condiciones de Kuhn-Tucker**, resulta no ser ni necesario ni suficiente para la obtención de un máximo. Sin embargo, las excepciones son tremendamente inusuales y pueden ser identificadas por no satisfacer la siguiente condición:

$$dg(x^*) \leq 0$$

De manera que si esto se cumple, las condiciones son necesarias. Si además la función objetivo es cóncava y las restricciones convexas, entonces son también suficientes.

1.3 Estática comparativa

El problema que abordamos a continuación es preguntarnos qué ocurre tanto con el punto óptimo como con el valor maximizado del objetivo cuando alguno de los parámetros del objetivo se modifica.

En efecto, sea

$$\begin{pmatrix} x_1^*(a) \\ \vdots \\ x_n^*(a) \end{pmatrix} = \arg \max f(x_1, \dots, x_n; a)$$

y el óptimo $y^* = f(x_1^*, \dots, x_n^*; a) = \max_{\mathbf{x} \in \mathbb{R}^n} f(x_1, \dots, x_n; a)$, vale decir, el valor maximizado de f para un nivel dado de a .

Nos preguntamos:

1. ¿Cómo cambian las variables óptimas al cambiar el parámetro?
2. ¿Cómo cambia el nivel del objetivo alcanzado?

La primera pregunta es lo que tradicionalmente se entiende por estática comparativa, y se centra en el signo (y ocasionalmente magnitud) de funciones de la forma:

$$\frac{\partial x_i^*(a)}{\partial a}$$

La segunda pregunta se refiere al máximo. Sobre el particular, usaremos intensivamente el siguiente resultado:

Teorema 1 (de la envolvente)

$$\frac{\partial y^*}{\partial a} = \frac{\partial f}{\partial a}$$

En efecto,

$$\frac{\partial y^*}{\partial a} = \sum_{i=1}^n \frac{\partial f(x_1^*, \dots, x_n^*; a)}{\partial x_i^*} \frac{\partial x_i^*}{\partial a} + \frac{\partial f(x_1^*, \dots, x_n^*; a)}{\partial a}$$

pero por condiciones de primer orden,

$$\begin{aligned} \frac{\partial f(x_1^*, \dots, x_n^*; a)}{\partial x_i^*} &= 0 \quad \forall i = 1, \dots, n \\ \Rightarrow \frac{\partial y^*}{\partial a} &= \frac{\partial f(x_1^*, \dots, x_n^*; a)}{\partial a} \end{aligned}$$

Este resultado es tremendamente importante porque permite simplificar notoriamente el análisis de las características del óptimo. En particular, nos dice que todos los efectos secundarios que el cambio en el parámetro provoca sobre la elección del óptimo son cero (puesto que de lo contrario no nos encontraríamos en el óptimo en primera instancia). Observe que ya usamos previamente esta idea para interpretar el multiplicador lagrangeano.