Resultados de optimización

Ricard Torres

August 8, 2007

Contents

1 Conceptos básicos. Optimización sin restricciones

1.1 Definiciones básicas. Existencia de soluciones

Distinguiremos en un problema de optimización tres elementos

1. El **conjunto factible** o **conjunto de oportunidades**, que denotaremos aquí con la letra X.
2. La **función objetivo**, una función definida sobre X que toma valores numéricos, $f : X \to \mathbb{R}$. La función f es un índice que asociamos a cada elemento de X.
3. Un criterio de decisión: buscaremos aquel elemento o elementos de X que tienen asociado un índice mayor (problemas de **maximización**) o bien un índice menor (problemas de **minimización**).

Simbólicamente, representaremos los problemas de maximización y de minimización como

$$\begin{aligned} \max_x \;& f(x) \\ \text{s.a.} \;& x \in X \end{aligned} \qquad \text{y} \qquad \begin{aligned} \min_x \;& f(x) \\ \text{s.a.} \;& x \in X \end{aligned}$$

donde las letras "s.a." significan "sujeto a". Debajo del símbolo **max** o **min** escribimos la variable x, que es la variable con respecto a la cual estamos optimizando, y que denominamos **variable de decisión** o **variable de control**.

Deberemos tener en cuenta que un problema de optimización, por sencillo que sea, no necesariamente tiene solución.

Ejemplo 1.1. Consideremos el problema

$$\begin{aligned} \max_x \;& x \\ \text{s.a.} \;& x \in \mathbb{N} \end{aligned}$$

donde $\mathbb{N}$ son los números naturales. Queremos hallar el mayor de todos los números naturales $\{1, 2, 3, 4, \ldots\}$, y este problema no tiene solución. ⊛

Definición 1.1 (Solución de un problema de maximización). Decimos que el problema

$$\begin{aligned} \max_x \;& f(x) \\ \text{s.a.} \;& x \in X \end{aligned}$$

tiene solución si hay un cierto elemento x^* en el conjunto X tal que se cumple que $f(x^*) \geq f(x)$, para todo x de X. En este caso decimos que x^* es un **maximizador** del problema, y que $f(x^*)$ es el **máximo** o **valor máximo** del problema.

Definición 1.2 (Solución de un problema de minimización). Decimos que el problema

$$\begin{aligned} \min_x \;& f(x) \\ \text{s.a.} \;& x \in X \end{aligned}$$

tiene solución si hay un cierto elemento x^* en el conjunto X tal que se cumple que $f(x^*) \leq f(x)$, para todo x de X. En este caso decimos que x^* es un **minimizador** del problema, y que $f(x^*)$ es el **mínimo** o **valor mínimo** del problema.

El hecho de que un problema de maximización tenga o no solución, depende de si el **recorrido** (la imagen) de la función f tiene un elemento máximo o no. Recordemos que, dada una función $f : X \to \mathbb{R}$, su recorrido (o imagen) está definido como

$$f(X) := \{c \in \mathbb{R} : \text{ para algún } x \in X, \ f(x) = c\}.$$

Lema 1.1. (i) El problema

$$\begin{aligned} \max_{x} \quad & f(x) \\ \text{s.a.} \quad & x \in X \end{aligned}$$

tiene solución si, y sólo si, el conjunto $f(X)$ (el recorrido de f) tiene un elemento máximo.

(ii) El problema

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.a.} \quad & x \in X \end{aligned}$$

tiene solución si, y sólo si, el conjunto $f(X)$ (el recorrido de f) tiene un elemento mínimo.

Recordaremos brevemente que, si A es un subconjunto de $\mathbb{R}$, decimos que a^* es el **elemento máximo** de A si $a^* \in A$ y, para cada $a \in A$, se cumple $a^* \geq a$. Análogamente, decimos que a^* es el **elemento mínimo** de A si $a^* \in A$ y, para cada $a \in A$, se cumple $a^* \leq a$. Una condición necesaria, pero no suficiente, para que un conjunto $A \subset \mathbb{R}$ tenga un máximo, es que A esté acotado superiormente. Para ver que la condición no es suficiente, basta considerar el siguiente ejemplo: el intervalo semiabierto $[0, 1)$ está acotado superiormente pero no tiene un máximo, ya que el número 1 no está en el intervalo, pero cualquier número positivo menor que 1 sí que está (por ejemplo 0.9999999 con tantos nueves como querramos). Una condición suficiente, pero no necesaria, para que A tenga un elemento máximo, y también un elemento mínimo, es que A sea **compacto**, es decir, acotado (superior e inferiormente) y cerrado.[1] Para ver que la condición no es necesaria, basta considerar el siguiente ejemplo: el conjunto $A = [0, 1) \cup 2$ (es decir, el conjunto formado por el intervalo semiabierto $[0, 1)$ y el número 2), tiene un máximo (el número 2), pero no es compacto, ya que no es cerrado.

El problema de dar condiciones que garanticen la existencia de una solución a un problema de maximización, es similar como hemos visto al problema de asegurar la existencia de elemento máximo en un conjunto de números reales: es fácil describir condiciones *suficientes* para que la solución exista, pero hay muchos casos en que una solución existe aunque aquellas condiciones no sean satisfechas.

[1] Un subconjunto A de $\mathbb{R}^n$ es **cerrado** si contiene los puntos límite de todas las secuencias que se pueden formar con sus elementos. El intervalo $[0, 1)$ no es cerrado, puesto que 1 no pertenece al mismo pero es el límite de la secuencia $1 - (1/n)$, para $n > 1$, que está enteramente contenida en dicho intervalo.

Teorema 1.1 (Teorema de Weierstrass). Si el conjunto factible X es acotado y cerrado (compacto) y la función objetivo f es continua, entonces tanto el problema de maximizar f sobre X como el de minimizarla tienen solución.

La prueba del teorema es sencilla: las condiciones impuestas garantizan que el recorrido $f(X)$ es un subconjunto compacto de $\mathbb{R}$, que siempre tiene un máximo y un mínimo. La utilidad del teorema es relativa, puesto que nos dice que existe una solución, pero no cómo hallarla.

1.2 Diferenciación y optimización. Condiciones de primer orden

Dado un punto $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, definimos el **rectángulo abierto de radio** $\varepsilon > 0$ **alrededor de** x como el conjunto:

$$\begin{aligned}(x_1 - \varepsilon, x_1 + \varepsilon) \; \times (x_2 - \varepsilon, x_2 + \varepsilon) \times \cdots (x_n - \varepsilon, x_n + \varepsilon) = \\ \{y \in \mathbb{R}^n \; : \; x_i - \varepsilon < y_i < x_i + \varepsilon, \; i = 1, 2, \dots, n\}\end{aligned}$$

Definición 1.3 (Entornos). Dado $x \in \mathbb{R}^n$, decimos que un conjunto E es un **entorno (o vecindario) de** x si existe algún $\varepsilon > 0$ tal que E contiene al rectángulo abierto de radio ε alrededor de x. Si X es un subconjunto de $\mathbb{R}^n$ y $x \in X$, entonces un **entorno de** x **relativo a** X es la intersección entre X y un entorno de x en $\mathbb{R}^n$.

Definición 1.4 (Máximos y mínimos locales). Sea $X \subset \mathbb{R}^n$ y $f : X \to \mathbb{R}$.

(i) Decimos que f alcanza un **máximo local** en un punto $x^* \in X$ si hay un cierto entorno de x^* relativo a X tal que $f(x^*) \geq f(x)$, para todo punto x de aquel entorno. Si la desigualdad es estricta para todo punto del entorno distinto de x^*, entonces decimos que el máximo local es **estricto**.

(ii) Decimos que f alcanza un **mínimo local** en un punto $x^* \in X$ si hay un cierto entorno de x^* relativo a X tal que $f(x^*) \leq f(x)$, para todo punto x de aquel entorno. Si la desigualdad es estricta para todo punto del entorno distinto de x^*, entonces decimos que el mínimo local es **estricto**.

Para distinguir entre maximizadores o minimizadores locales y los que solucionan un determinado problema de optimización, solemos añadir a estos últimos la equiqueta de **globales**.

Dado un subconjunto X de $\mathbb{R}^n$, decimos que un punto $x \in X$ es **interior** a X si existe un entorno de x que está enteramente contenido en el conjunto. Un subconjunto es **abierto** si, y sólo si, todos sus puntos son interiores. Un subconjunto es **cerrado** si, y sólo si, su complementario es un conjunto abierto.

Teorema 1.2 (Regla de Fermat). Sea $f : X \to \mathbb{R}$ una función diferenciable. Si x^* es un punto interior del conjunto X y f alcanza un máximo o un mínimo local en x^*, entonces todas las derivadas parciales de f evaluadas en x^* son iguales a cero.

En el caso univariante, el motivo es el conocido de que, si no fuera así, la función estaría creciendo o decreciendo estrictamente alrededor de x. En el caso multivariante, el mismo argumento se aplica a las secciones longitudinales de la función.

Definición 1.5 (Condiciones de primer orden). Denominamos **condiciones de primer orden** en un problema de optimización sin restricciones al sistema de ecuaciones, en término de las derivadas parciales de primer orden, que describe los puntos críticos de la función objetivo, es decir, el que resulta de igualar a cero todas las derivadas parciales de la función.

1.3 Condiciones de segundo orden

Las condiciones de segundo orden del problema de optimización son aquellas que involucran derivadas parciales de segundo orden, y que, en algunos casos, nos permiten determinar la naturaleza de los puntos críticos que hemos hallado mediante las condiciones de primer orden.

Teorema 1.3 (Condiciones Suficientes de Segundo Orden). Sea $X \subset \mathbb{R}^n$ un conjunto abierto, y sea $f : X \to \mathbb{R}$ una función que tiene derivadas parciales de segundo orden que son funciones continuas. Supongamos que $\bar{x} \in \mathbb{R}^n$ es un punto crítico, es decir que, para cada i, $1 \leq i \leq n$, $D_i f(\bar{x}) = 0$. Tenemos que

(i) Si la matriz $D^2 f(\bar{x})$ es definida positiva, entonces $\bar{x}$ es un minimizador local (estricto) de la función f.

(ii) Si la matriz $D^2 f(\bar{x})$ es definida negativa, entonces $\bar{x}$ es un maximizador local (estricto) de la función f.

(iii) Si la matriz $D^2 f(\bar{x})$ es indefinida, entonces $\bar{x}$ es un punto de silla de la función f.

Observación: Ninguna de las implicaciones del anterior teorema es cierta si invertimos su sentido (e.g., no es cierto que un maximizador local estricto tenga siempre asociada una matriz definida negativa).

Cuando el criterio anterior no sea decisorio, como mínimo tal vez podamos descartar alguna posibilidad usando las condiciones *necesarias* de segundo orden.

Teorema 1.4 (Condiciones Necesarias de Segundo Orden). Sea $X \subset \mathbb{R}^n$ un conjunto abierto, y sea $f : X \to \mathbb{R}$ una función que tiene derivadas parciales de segundo orden que son funciones continuas. Entonces:

(i) Si $\bar{x}$ es un minimizador local (no necesariamente estricto), entonces $\bar{x}$ es un punto crítico y la matriz hessiana $D^2 f(\bar{x})$ es semidefinida positiva.

(ii) Si $\bar{x}$ es un maximizador local (no necesariamente estricto), entonces $\bar{x}$ es un punto crítico y la matriz hessiana $D^2 f(\bar{x})$ es semidefinida negativa.

Observación: Ninguna de las implicaciones del anterior teorema es cierta si invertimos su sentido.

2 Concavidad y convexidad de funciones. Criterios de globalidad

2.1 Conceptos básicos

Definición 2.1 (Conjunto convexo). Sea X un subconjunto de $\mathbb{R}^n$. Decimos que X es un *conjunto convexo* si el segmento que une dos puntos cualesquiera del conjunto está enteramente contenido en el mismo. Es decir, X es convexo si, dados dos puntos cualesquiera x' y x'' de X, y dada cualquier $\lambda \in (0, 1)$, el punto $\lambda x' + (1 - \lambda)x''$ también pertenece a X.

Es sencillo probar que la intersección de conjuntos convexos resulta en un conjunto que también es convexo.

Definición 2.2 (Funciones cóncavas y convexas). Sea X un subconjunto convexo de $\mathbb{R}^n$, y $f : X \to \mathbb{R}$ una función.

(i) Decimos que f es una función cóncava si, y sólo si, dados dos puntos cualesquiera x' y x'' de X, y dada cualquier $\lambda \in (0, 1)$, se cumple la siguiente desigualdad:

$$f\big(\lambda x' + (1 - \lambda)x''\big) \geq \lambda\, f(x') + (1 - \lambda)\, f(x'')$$

Si la desigualdad es estricta cuando $x' \neq x''$, entonces decimos que f es estrictamente cóncava.

(ii) Decimos que f es una función convexa si, y sólo si, dados dos puntos cualesquiera x' y x'' de X, y dada cualquier $\lambda \in (0, 1)$, se cumple la siguiente desigualdad

$$f\big(\lambda x' + (1 - \lambda)x''\big) \leq \lambda\, f(x') + (1 - \lambda)\, f(x'')$$

Si la desigualdad es estricta cuando $x' \neq x''$, entonces decimos que f es estrictamente convexa.

Supongamos que f es una función diferenciable. Denominamos $\nabla f(\bar{x})$ el *vector gradiente* de la función f evaluado en el punto $\bar{x}$, es decir, el vector compuesto por las derivadas parciales de la función f

evaluadas en el punto $\bar{x}$. Mediante un punto '·' designaremos el producto escalar entre vectores. La tangente al gráfico de f en el punto $\big(\bar{x}, f(\bar{x})\big)$ es el gráfico de la función:

$$g(x) = f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x})$$

Geométricamente, la definición de función cóncava implica que toda tangente al gráfico de la función queda enteramente por encima de este último. Formalmente, esto significa que, para cada $\bar{x} \in X$ y cada $x \in X$, se cumple que

$$f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) \geq f(x)$$

Proposición 2.1 (Concavidad, convexidad y derivadas primeras). Sea $X \subset \mathbb{R}^n$ un conjunto convexo y $f : X \to \mathbb{R}$ una función diferenciable.

(i) f es una función cóncava si, y sólo si, dados dos puntos cualesquiera x' y x'', se cumple que

$$f(x') + \nabla f(x') \cdot (x'' - x') \geq f(x'')$$

(ii) f es una función estrictamente cóncava si, y sólo si, dados dos puntos distintos x' y x'', se cumple que

$$f(x') + \nabla f(x') \cdot (x'' - x') > f(x'')$$

(iii) f es una función convexa si, y sólo si, dados dos puntos cualesquiera x' y x'', se cumple que

$$f(x') + \nabla f(x') \cdot (x'' - x') \leq f(x'')$$

(iv) f es una función estrictamente convexa si, y sólo si, dados dos puntos distintos x' y x'', se cumple que

$$f(x') + \nabla f(x') \cdot (x'' - x') < f(x'').$$

Decimos que una función es *diferenciable con continuidad* si sus derivadas parciales son funciones continuas. Análogamente, una función es *diferenciables dos veces con continuidad* si sus derivadas parciales segundas son funcions continuas.

Recordemos que un subconjunto de $\mathbb{R}^n$ es **abierto** si todos sus puntos son interiores.

Proposición 2.2 (Concavidad, convexidad y derivadas segundas). Sea $X \subset \mathbb{R}^n$ un conjunto abierto convexo, y $f : X \to \mathbb{R}$ una función diferenciable dos veces con continuidad.

(i) f es una función cóncava si, y sólo si, para cada $x \in X$ la matriz hessiana $D^2 f(x)$ es semidefinida negativa.

(ii) f es una función convexa si, y sólo si, para cada $x \in X$ la matriz hessiana $D^2 f(x)$ es semidefinida positiva.

Teorema 2.1 (Concavidad o convexidad estricta y derivadas segundas). Sea $X \subset \mathbb{R}^n$ un conjunto abierto convexo, y $f : X \to \mathbb{R}$ una función diferenciable dos veces con continuidad.

(i) Si para cada $x \in X$ la matriz hessiana $D^2 f(x)$ es definida negativa, entonces f es estrictamente cóncava.

(ii) Si para cada $x \in X$ la matriz hessiana $D^2 f(x)$ es definida positiva, entonces f es estrictamente convexa.

2.2 Concavidad, convexidad y optimización

Teorema 2.2 (Teorema local–global para funciones cóncavas). Sea $X \subset \mathbb{R}^n$ un conjunto convexo y $f : X \to \mathbb{R}$ una función cóncava, y supongamos que el punto $x^* \in X$ es un maximizador local de f. Entonces x^* es también un maximizador global.

Teorema 2.3 (Teorema local–global para funciones convexas). Sea $X \subset \mathbb{R}^n$ un conjunto convexo y $f : X \to \mathbb{R}$ una función convexa, y supongamos que el punto $x^* \in X$ es un minimizador local de f. Entonces x^* es también un minimizador global.

Observaciones:

- En un problema de maximización, nos interesará que la función objetivo sea cóncava, ya que en este caso sabremos que los máximos locales lo son también en un sentido global.
- En un problema de minimización, nos interesará que la función objetivo sea convexa, ya que en este caso sabremos que los mínimos locales lo son también en un sentido global.
- Notemos que el maximizador o minimizador local al cual se refiere el teorema no tiene por qué ser un punto interior. De hecho, el único requerimiento sobre el dominio es que se trate de un conjunto convexo. Esto permite derivar resultados de globalidad en todo tipo de problemas de optimización, tengan o no restricciones.
- Si la función es estrictamente cóncava (convexa), entonces el maximizador (minimizador) a que se refiere el teorema es único.

DEMOSTRACIÓN: Demostraremos el resultado en el caso de problemas de maximización y concavidad de la función f. Para mostrar que x^* es un maximizador global, debemos probar que, para todo $x \in X$, $f(x^*) \geq f(x)$. Sea x un punto arbitrario de X. Sabemos que, para cualquier $\lambda \in (0,1)$, $(1-\lambda)\,x^* + \lambda\,x \in X$. Notemos que, cuando λ está muy cercana a 0, el punto $(1-\lambda)\,x^* + \lambda\,x$ está muy cercano a x^*. Por definición de maximizador local, habrá una cierta $\bar{\lambda}$, con $1 > \bar{\lambda} > 0$ suficientemente cercana a 0, tal que para toda $\bar{\lambda} > \lambda > 0$ se cumple:

$$f(x^*) \geq f\big((1-\lambda)\,x^* + \lambda\,x\big)$$

y por concavidad

$$f\big((1-\lambda)\,x^* + \lambda\,x\big) \geq (1-\lambda)\,f(x^*) + \lambda\,f(x)$$

Combinando ambas desigualdades, obtenemos

$$f(x^*) \geq (1-\lambda)\,f(x^*) + \lambda\,f(x)$$

y, en consecuencia,

$$\lambda\,f(x^*) \geq \lambda\,f(x)$$

Finalmente, como λ es positiva, podemos eliminarla de la desigualdad y obtenemos $f(x^*) \geq f(x)$. Como x era un punto arbitrario, esto demuestra que x^* es un maximizador global de la función f.

El resultado paralelo para funciones convexas y problemas de minimización se demuestra en la misma forma. □

Nota: Debemos tener presente que es esencial que el dominio de la función sea un conjunto convexo.

En base a la caracterización de funciones cóncavas y convexas diferenciables, podemos derivar inmediatamente otro criterio sumamente útil.

Teorema 2.4. Sea $X \subset \mathbb{R}^n$ un conjunto convexo y $f : X \to \mathbb{R}$ una función diferenciable.

(i) Si f es una función cóncava, entonces cualquier punto crítico (donde todas las derivadas parciales se anulan) es un maximizador global.

(ii) Si f es una función convexa, entonces cualquier punto crítico (donde todas las derivadas parciales se anulan) es un minimizador global.

3 Optimización con restricciones de igualdad

3.1 Condiciones de primer orden: el método de los multiplicadores de Lagrange

En problemas con restricciones de igualdad, dichas restricciones establecen una relación de dependencia entre las variables. De hecho, cuando las restricciones son lineales y no hay problemas de singularidad, se pueden expresar algunas variables en términos de las demás y convertir el problema en uno de optimización libre con menos variables. El método de los multiplicadores de Lagrange adopta un enfoque distinto: ignoramos la dependencia entre las variables y optimizamos libremente una función con *todas* las variables, pero en la cual a la función objetivo se han añadido unos términos de penalización que aseguran que las restricciones acaben siendo satisfechas.

Teorema 3.1 (Regla de los Multiplicadores de Lagrange). Sean $f, g_1, g_2, \ldots, g_m$, funciones definidas en un subconjunto abierto de $\mathbb{R}^n$ y que toman valores en $\mathbb{R}$, con $m \leq n$. Supongamos que todas las funciones son diferenciables con continuidad (sus derivadas parciales son funciones continuas).

Sea x^* un maximizador o minimizador local de f sobre el conjunto de puntos que satisfacen

$$\begin{aligned} g_1(x_1, x_2, \ldots, x_n) &= 0 \\ g_2(x_1, x_2, \ldots, x_n) &= 0 \\ \cdots\cdots\cdots\cdots \\ g_m(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}$$

Si los gradientes de g_1, g_2, ..., g_m evaluados en x^* son linealmente independientes, entonces hay números reales $\lambda_1, \lambda_2, \ldots, \lambda_m$, que cumplen

$$\nabla f(x^*) \;=\; \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) + \cdots + \lambda_m \nabla g_m(x^*)$$

Una manera conveniente de obtener las condiciones que menciona el teorema anterior es definiendo una nueva función: el **lagrangiano** o función de Lagrange.

Definición 3.1 (Lagrangiano). Dadas funciones f, g_1, g_2, ..., g_m, de la misma forma que en el teorema anterior, definimos el **lagrangiano** o función de Lagrange como la función de $n+m$ argumentos:

$$L(x, \lambda) = f(x) - \lambda_1 g_1(x) - \lambda_2 g_2(x) - \cdots - \lambda_m g_m(x),$$

donde $x = (x_1, x_2, \ldots, x_n)$ y $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$.

La interpretación intuitiva del lagrangiano es que se trata de una nueva función objetivo, que nos permite olvidarnos de que las variables han de obedecer las restricciones. Construimos el lagrangiano añadiendo a la función objetivo original un *término de penalización* para cada restricción, de tal forma que los términos adicionales de penalización sólo tienen efecto cuando las respectivas restricciones no se cumplen. Para aplicar la regla de los multiplicadores de Lagrange, definimos en primer lugar el lagrangiano y después igualamos a cero las derivadas parciales del mismo con respecto a cada una de las variables originales. Finalmente, añadimos las restricciones para obtener un sistema con tantas ecuaciones como incógnitas.

Ejemplo 3.1. El lector estará familiarizado con funciones de una sola variable, digamos $h(x)$, donde las soluciones de $h'(x) = 0$ no son maximizadores ni minimizadores, o bien lo son sólo localmente. Para ver que en los problemas con restricciones de igualdad podemos encontrarnos con extremos locales que no lo son en un sentido global, o con puntos críticos que no son ni maximizadores ni minimizadores, nos basta con observar que los dos problemas siguientes son equivalentes:

$$\begin{array}{rl} \max\limits_{x} & h(x) \\ \text{s.a.} & x \in \mathbb{R} \end{array} \qquad \Longleftrightarrow \qquad \begin{array}{rl} \max\limits_{(x,y)} & y \\ \text{s.a.} & y = h(x) \end{array}$$

Es sencillo comprobar que, si aplicamos el método de Lagrange al segundo problema, las condiciones de primer orden se reducen a $h'(x) = 0$.

Si en el problema anterior tomamos $h(x) = x^3$, encontraremos un punto crítico que no es maximizador ni minimizador. Por otro lado, tomando $h(x) = 5x^2 - 3x^4 + \frac{1}{3}x^6$ nos encontraremos con maximizadores y minimizadores locales, pero no globales. ⊛

3.2 Regularidad

Antes hemos visto que, para aplicar el método de Lagrange, necesitamos que los gradientes de las restricciones evaluados en el maximizador o minimizador sean linealmente independientes. Cuando esta condición no se cumple, los gradientes de las restricciones no describen correctamente el plano tangente al conjunto factible, y las cosas fallan: pueden existir soluciones al problema de optimización que no aparecen cuando aplicamos el método de Lagrange.

Ejemplo 3.2. Consideremos el problema

$$\begin{aligned} \max_{(x,y)} \ & -x^2 + y \\ \text{s.a.} \ & y^3 = 0 \end{aligned}$$

Notemos que la restricción es equivalente a $y = 0$, y substituyendo en la función objetivo vemos que la solución es $x^* = y^* = 0$. Aplicando el método de Lagrange nos encontramos con las ecuaciones

$$\begin{aligned} -2\,x &= 0 \\ 1 - 3\,\lambda\, y^2 &= 0 \\ y^3 &= 0 \end{aligned}$$

La tercera ecuación implica $y = 0$, y substituyendo en la segunda nos quedamos con $1 = 0$, por lo cual no hay ningún punto que satisfaga estas condiciones. Recapitulando: nos hemos encontrado con un problema que sabemos que tiene solución, y sin embargo el método de Lagrange nos dice que no hay ninguna solución.

Notemos que el gradiente de la restricción es $\nabla g(x, y) = (0, 3y^2)$, y en el punto que sabemos que es el maximizador del problema tenemos $\nabla g(0, 0) = (0, 0)$, que es linealmente dependiente (recordemos que el vector de ceros no forma nunca parte de un conjunto de vectores linealmente independientes).

Geométricamente, como la restricción es el eje de las abcisas, el plano tangente a la restricción en cualquier punto sigue siendo el eje de las abcisas. Pero como el gradiente de g es el vector 0, el conjunto de vectores ortogonales al gradiente de g no es el plano tangente, sino todo el espacio $\mathbb{R}^2$. ⊛

Definición 3.2 (Regularidad). Decimos que un punto factible es **regular** si los vectores gradientes de las restricciones son linealmente independientes cuando los evaluamos en dicho punto.

El teorema que hemos visto anteriormente afirma que el método de los multiplicadores de Lagrange detecta cualquier maximizador o minimizador local en un problema con restricciones de igualdad, *siempre que éstos sean puntos regulares*.

Notemos que la regularidad es una condición suficiente, pero no necesaria, para la validez de la regla de los multiplicadores de Lagrange. Un análisis geométrico nos puede ayudar a ver en qué casos el método de Lagrange detectará maximizadores o minimizadores donde falla la condición de regularidad, o dejará de hacerlo.

Ejemplo 3.3. Consideremos un problema con dos variables y la restricción $g(x, y) = x^2 + y^3 = 0$. Como $\nabla g(x, y) = (2x, 3y^2)$, todos los puntos factibles son regulares con excepción del punto $(0, 0)$. Dada una función f que tiene un maximizador o minimizador local sobre esta superficie en $(0, 0)$, queremos saber si el método de Lagrange tendrá este punto como solución o no. Todo se reduce a verificar si se cumple la "condición de tangencia" $\nabla f(0, 0) = \lambda\, \nabla g(0, 0)$. Como $\nabla g(x, y) = (0, 0)$, el método de Lagrange funcionará si $\nabla f(0, 0) = (0, 0)$, es decir, si $(0, 0)$ es un punto crítico (sin restricciones) de f. Por ejemplo, si $f(x, y) = y$ el método de Langrange no funcionará, pero sí que lo hará si $f(x, y) = x^2 + y^2$. ⊛

Como hemos visto, cuando hay una sola restricción, el método de Lagrange detectará un maximizador o minimizador no regular si, y sólo si, éste es un punto crítico de la función. Con más de una restricción, hay otros casos a tener en cuenta.

Ejemplo 3.4. Consideremos las dos restricciones

$$\begin{aligned} g_1(x,y,z) &= (x-1)^2 + y^2 = 1 \\ g_2(x,y,z) &= (x+1)^2 + y^2 = 1 \end{aligned}$$

Aquí tenemos $\nabla g_1(x,y,z) = \big(2\,(x-1), 2\,y, 0\big)$ y $\nabla g_2(x,y,z) = \big(2\,(x+1), 2\,y, 0\big)$. Como el conjunto factible es

$$X = \left\{(x,y,z) \in \mathbb{R}^3 : \, x = y = 0\right\},$$

todos los puntos factibles satisfacen

$$\begin{aligned} \nabla g_1(x,y,z) &= \big(-2,0,0\big) \\ \nabla g_2(x,y,z) &= \big(2,0,0\big) \end{aligned}$$

Es decir, no hay ningún punto factible que sea regular. La función $f(x,y,z) = x$ es constante sobre el conjunto factible, y por tanto tiene tanto un máximo como un mínimo global en cada punto factible; el método de Lagrange funcionará en este caso, ya que $\nabla f(x,y,z) = (1,0,0)$ es linealmente dependiente de cualquiera de los dos gradientes de las restricciones. En cambio, la función $f(x,y,z) = y$, que también es constante sobre el conjunto factible, tiene como gradiente $\nabla f(x,y,z) = (0,1,0)$, que no depende linealmente de los gradientes de las restricciones. ¿Cuál es la diferencia en los dos casos? La condición de tangencia exige que el plano tangente a la curva de nivel de la función objetivo en el punto crítico contenga el eje de las z, que es el conjunto factible. Sin embargo, la dependencia lineal de los gradientes de las restricciones hace que analíticamente solamente podamos detectar uno de los muchos planos tangentes que satisfacen dicha propiedad, que es el plano y–z. El método de Lagrange funcionará si, y sólo si, el gradiente de f en el punto crítico es ortogonal al plano y–z. ⊛

3.3 Criterios de globalidad

En el caso con restricciones de igualdad, el criterio de concavidad/convexidad sólo puede ser directamente aplicado cuando todas las restricciones son *ecuaciones lineales*, porque es esencialmente el único caso[2] en que el conjunto factible es convexo. En este caso, el teorema local–global nos permite inferir el carácter global de los maximizadores o minimizadores locales. Otra forma de aplicar criterios de concavidad/convexidad, consiste en analizar la curvatura del lagrangiano. Recordemos que, en un problema de optimización con restricciones de igualdad, el conjunto factible es el que forman los puntos que satisfacen todas las restricciones.

Teorema 3.2 (Concavidad o convexidad del lagrangiano). Dado un problema de optimización con restricciones de igualdad, sea X el conjunto factible, y supongamos que hay un conjunto convexo Y tal que $X \subseteq Y \subseteq \mathbb{R}^n$ y las funciones f, g_1, g_2, ..., g_m, están definidas y son diferenciables con continuidad en Y. Sea (x^*, λ^*) un punto que soluciona las condiciones de primer orden del método

[2] Exceptuando restricciones que tienen más de una representación. Por ejemplo, la restricción lineal $y = 0$ también puede ser escrita en forma no lineal como $y^3 = 0$. La anomalia de esta representación queda patente en el ejemplo anterior donde debido a ella el método de los multiplicadores de Lagrange no es aplicable.

de Lagrange, donde $x = (x_1, x_2, \ldots, x_n)$ y $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$. Definimos la función $h : Y \to \mathbb{R}$ per $h(x) = L(x, \lambda^*)$; es decir, h es el lagrangiano cuando mantenemos fijos los multiplicadores en los valores λ^* y tomamos el vector x como variables. Entonces tenemos:

(i) Si h es una función cóncava sobre Y, x^* es un maximizador global del problema con restricciones de igualdad.

(ii) Si h es una función convexa sobre Y, x^* es un minimizador global del problema con restricciones de igualdad.

DEMOSTRACIÓN: Supongamos que h es una función cóncava sobre Y. Las condiciones de tangencia implican que, para cada $j = 1, 2, \ldots, n$ se cumple

$$\frac{\partial L}{\partial x_j}(x^*, \lambda^*) = \frac{\partial h}{\partial x_j}(x^*) = 0,$$

es decir, todas las derivadas parciales de h son 0 cuando las evaluamos en el vector x^*. Como h es una función cóncava, esto implica que x^* es un maximizador global de h sobre Y, es decir, $h(x^*) \geq h(x)$, para todo $x \in Y$ (satisfaga o no las restricciones de igualdad). Tenemos que, para todo $x \in Y$:

$$h(x^*) \geq h(x) \iff f(x^*) - \sum_{i=1}^{m} \lambda_i^* \, g_i(x^*) \geq f(x) - \sum_{i=1}^{m} \lambda_i^* \, g_i(x)$$

Como x^* satisface las restricciones de igualdad, tenemos que, para cada $i = 1, 2, \ldots, m$, $g_i(x^*) = 0$ y, por tanto, para cualquier $x \in Y$ se cumple:

$$f(x^*) \geq f(x) - \sum_{i=1}^{m} \lambda_i^* \, g_i(x)$$

En particular, si $x \in X$ también se cumple que, para cada $i = 1, 2, \ldots, m$, $g_i(x) = 0$. Es decir, para todo $x \in X$,

$$f(x^*) \geq f(x)$$

Por tanto, x^* es un maximizador global del problema con restricciones de igualdad. El caso de minimización se demuestra en forma similar. □

Ejemplo 3.5. Sea $f(x, y) = x\, y$ y $g(x, y) = x^2 + y^2 - 2$, y consideremos el problema de hallar los máximos o mínimos de f sobre la superficie definida por $g(x, y) = x^2 + y^2 - 2 = 0$. Notemos que, en este caso, podem tomar $Y = \mathbb{R}^2$. El lagrangiano es $L(x, y, \lambda) = x\, y - \lambda\, (x^2 + y^2 - 2)$, y las condiciones de primer orden son

$$\begin{aligned} \frac{\partial L}{\partial x} = y - 2\,\lambda\, x &= 0 \\ \frac{\partial L}{\partial y} = x - 2\,\lambda\, y &= 0 \\ x^2 + y^2 &= 2 \end{aligned}$$

Este sistema de ecuaciones tiene cuatro soluciones, donde (x, y, λ) adoptan los valores

$$(1, 1, 1/2), \quad (-1, -1, 1/2), \quad (-1, 1, -1/2) \quad \text{y} \quad (1, -1, -1/2)$$

Tomemos la primera solución: $(x, y, \lambda) = (1, 1, 1/2)$. Fijamos $\lambda = 1/2$ y dejamos (x, y) como variables del lagrangiano

$$h(x, y) = L(x, y, \frac{1}{2}) = x\,y - \frac{1}{2}\,\left(x^2 + y^2 - 2\right)$$

Las derivadas de primer orden de h son

$$\frac{\partial h}{\partial x} = y - x \qquad \text{y} \qquad \frac{\partial h}{\partial y} = x - y$$

En particular, como sabemos por las condiciones de tangencia, se cumple

$$\frac{\partial h}{\partial x}(1, 1) = 0 \qquad \text{y} \qquad \frac{\partial h}{\partial y}(1, 1) = 0$$

La matriz hessiana de h es

$$D^2h(x, y) = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

que es semidefinida negativa. Como eso es cierto para cada $(x, y) \in \mathbb{R}^2$, sabemos que h es una función cóncava sobre $\mathbb{R}^2$. Como h es cóncava y sus derivadas parciales son cero en el punto $(1, 1)$, sabemos que este punto es un maximizador global de h. El teorema que acabamos de ver nos asegura que $(1, 1)$ es también un maximizador global de la función $f(x, y) = x\,y$ sobre la superficie definida por $g(x, y) = x^2 + y^2 - 2 = 0$.

Un análisis similar para cada uno de los demás puntos, nos lleva a concluir que tanto $(1, 1)$ como $(-1, -1)$ son maximizadores globales, mientras que $(1, -1)$ y $(-1, 1)$ son minimizadores globales. ⊛

Ejemplo 3.6. Supongamos que $f(x, y) = -x^2 + y$ y $g(x, y) = x^2 + y^2 - 1 = 0$. Dado que todas las funciones son polinómicas, podemos tomar $Y = \mathbb{R}^2$. Las condiciones de primer orden dan cuatro puntos críticos:

$$\begin{aligned} (x^*, y^*, \lambda^*) &= (0, 1, \tfrac{1}{2}) \\ (x^*, y^*, \lambda^*) &= (0, -1, -\tfrac{1}{2}) \\ (x^*, y^*, \lambda^*) &= (\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1) \\ (x^*, y^*, \lambda^*) &= (-\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1) \end{aligned}$$

de los cuales los dos primeros son maximizadores locales y los dos últimos son minimizadores locales. El lagrangiano es $L(x, y, \lambda) = -x^2 + y - \lambda\,\left(x^2 + y^2 - 1\right)$.

Consideremos el primer punto $(x^*, y^*, \lambda^*) = (0, 1, \frac{1}{2})$. De donde $h(x, y) = -x^2 + y - \frac{1}{2}\,\left(x^2 + y^2 - 1\right)$. La matriz hessiana de h es:

$$D^2h(x, y) = \begin{pmatrix} -3 & 0 \\ 0 & -1 \end{pmatrix}$$

Como $D^2h(x, y)$ es semidefinida negativa en cada $(x, y) \in \mathbb{R}^2$, h es una función cóncava sobre $\mathbb{R}^2$. El teorema de globalidad implica que $(x^*, y^*) = (0, 1)$ es un maximizador global.

Tomemos ahora el punto $(x^*, y^*, \lambda^*) = (0, -1, -\frac{1}{2})$, de donde $h(x, y) = -x^2 + y + \frac{1}{2}\,\left(x^2 + y^2 - 1\right)$. La matriz hessiana de h es ahora:

$$D^2h(x, y) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Como $D^2h(x, y)$ es indefinida en todo punto, h no es cóncava ni convexa y el teorema no es aplicable. De hecho, observando el gráfico del problema, podemos ver que este punto es un maximizador local, pero no global.

Finalmente, tomando $(\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1)$ o bien $(-\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1)$, como tienen la mateixa $\lambda = -1$, les corresponde a ambos la misma función $h(x, y) = -x^2 + y + \left(x^2 + y^2 - 1\right)$. La matriz hessiana de h es:

$$D^2 h(x, y) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

Como $D^2 h(x, y)$ es semidefinida positiva en cada $(x, y) \in \mathbb{R}^2$, h es una función convexa sobre $\mathbb{R}^2$. El teorema de globalidad implica que los dos puntos son minimizadores globales. ⊛

Si a una función le añadimos términos lineales, sus derivadas segundas no varían, y por tanto su concavidad o convexidad es la misma. Como el lagrangiano resulta de añadir las funciones que definen las restricciones a la función objetivo, una consecuencia del anterior teorema es:

Corolario 3.1 (Globalidad y restricciones lineales). Supongamos que, en un problema con restricciones de igualdad, todas las restricciones son ecuaciones lineales. El conjunto factible X es un subespacio lineal, y por tanto un conjunto convexo. Entonces tendremos:

(i) Si f es una función cóncava sobre X, toda solución de las condiciones de primer orden del método de Lagrange es un maximizador global.

(ii) Si f es una función convexa sobre X, toda solución de las condiciones de primer orden del método de Lagrange es un minimizador global.

Observación: Notemos que este resultado es más sencillo de aplicar que el teorema local–global. La diferencia estriba en que para aplicar este teorema necesitamos primero conocer que el candidato es un maximizador o minimizador local, mientras que el corolario que acabamos de ver solamente requiere que el punto resuelva las condiciones del método de Lagrange.

Ejemplo 3.7. En el problema

$$\begin{aligned} \max_{(x,y)} \quad & x\, y \\ \text{s.a.} \quad & x + y = 1 \end{aligned}$$

la única solución de las condiciones de primer orden del método de Langrange es el punto $(x^*, y^*, \lambda) = (1/2, 1/2, 1/2)$. Claramente, la función objetivo no es cóncava sobre $\mathbb{R}^2$, pero sí lo es sobre el conjunto factible, por lo que el corolario anterior se aplica. ⊛

3.4 Condiciones de segundo orden

Denotemos mediante $D_x^2 L(x, \lambda)$ la submatriz hessiana formada por las derivadas parciales segundas del lagrangiano con respecto a las variables x y, dado cualquier vector $h \in \mathbb{R}^n$, sea $D_x^2 L(x, \lambda)\big(h, h\big)$ la forma cuadrática generada por aquella submatriz hessiana aplicada al vector h (escribimos el vector h dos veces porque hemos de pre- y post-multiplicar la matriz por el vector). Dado un cierto x factible y regular, el plano tangente al conjunto factible que pasa por el punto x viene dado por[3]

$$T(x) = \{h \in \mathbb{R}^n : \ \nabla g_i(x) \cdot h = 0, \ \text{para } 1 \leq i \leq m\}$$

[3] Si el punto no es regular, el plano tangente está contenido dentro del conjunto $T(x)$, pero puede no coincidir con el mismo.

Teorema 3.3 (Condiciones Suficientes de Segundo Orden). Supongamos que $m \leq n$ y que f, g_1, g_2, ..., g_m, son diferenciables dos veces con continuidad (sus derivadas parciales segundas son funciones continuas). Sea x^* un punto factible que satisface las condiciones de primer orden del método de Lagrange, y sea λ^* el correspondiente multiplicador de Lagrange. Entonces tenemos:

(i) *(Condición suficiente para maximización local)* Supongamos que $D_x^2 L(x^*, \lambda^*)\big(h, h\big) < 0$ para cada $h \in T(x^*)$ tal que $h \neq 0$. Entonces x^* es un maximizador local estricto.

(ii) *(Condición suficiente para minimización local)* Supongamos que $D_x^2 L(x^*, \lambda^*)\big(h, h\big) > 0$ para cada $h \in T(x^*)$ tal que $h \neq 0$. Entonces x^* es un minimizador local estricto.

Hay una regla, basada en equivalencias algebraicas, que permite comprobar estas condiciones de segundo orden mirando sólo una cierta matriz.

Supongamos que estamos analizando los puntos críticos de la función

$$f(x_1, x_2, \ldots, x_n)$$

sobre la superficie definida por

$$\begin{aligned} g_1(x_1, x_2, \ldots, x_n) &= 0 \\ g_2(x_1, x_2, \ldots, x_n) &= 0 \\ \cdots\cdots\cdots\cdots \\ g_m(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}$$

Escribamos el lagrangiano : $L(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_m)$. Dado r tal que $m + 1 \leq r \leq n$, definimos el siguiente determinante

$$\Delta_r(x, \lambda) = \begin{vmatrix} 0 & \cdots & 0 & \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_r} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_r} \\ \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_1} & \frac{\partial^2 L}{\partial x_1^2} & \cdots & \frac{\partial^2 L}{\partial x_1 \, \partial x_r} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_r} & \cdots & \frac{\partial g_m}{\partial x_r} & \frac{\partial^2 L}{\partial x_r \, \partial x_1} & \cdots & \frac{\partial^2 L}{\partial x_r^2} \end{vmatrix}$$

Proposición 3.1 (Método de la hessiana orlada). Supongamos que f, g_1, g_2, ..., g_m, son diferenciables dos veces con continuidad (sus derivadas parciales segundas son funciones continuas). Sea x^* un punto factible regular que satisface las condiciones de primer orden del método de Lagrange, y sea λ^* el correspondiente multiplicador de Lagrange. Entonces tenemos:

(i) *(Condición suficiente para la maximización local)* Si para cada $r = m + 1, m + 2, \ldots, n$ se cumple que $(-1)^r \, \Delta_r(x^*, \lambda^*) > 0$, entonces x^* es un maximizador local estricto.

(ii) *(Condición suficiente para la minimización local)* Si para cada $r = m+1, m+2, \ldots, n$ se cumple que $(-1)^m \, \Delta_r(v^*, \lambda^*) > 0$, entonces v^* es un minimizador local estricto.

Si no hubiera restricciones, haciendo $m = 0$ y $L = f$, las condiciones de segundo orden que acabamos de exponer se transforman en las condiciones en términos de menores preferentes que son conocidas en los problemas de optimización sin restricciones.

Teorema 3.4 (Condiciones Necesarias de Segundo Orden). Supongamos que $m \leq n$ y que f, g_1, $g_2, \ldots, g_m$, son diferenciables dos veces con continuidad (sus derivadas parciales segundas son funciones continuas).

(i) *(Condición necesaria para maximización local)* Si x^* es un maximizador local y es un punto regular, entonces existe λ^* tal que (x^*, λ^*) satisfacen las condiciones de primer orden del método de Lagrange, y $D_x^2 L(x^*, \lambda^*)\big(h, h\big) \leq 0$ para cada $h \in T(x^*)$.

(ii) *(Condición necesaria para minimización local)* Si x^* es un minimizador local y es un punto regular, entonces existe λ^* tal que (x^*, λ^*) satisfacen las condiciones de primer orden del método de Lagrange, y $D_x^2 L(x^*, \lambda^*)\big(h, h\big) \geq 0$ para cada $h \in T(x^*)$.

4 Optimización con restricciones de desigualdad

4.1 El método de Kuhn–Tucker

Definición 4.1 (Restricción efectiva). Decimos que una restricción es **efectiva** (o saturada, o activa) en un punto si éste la satisface con igualdad.

Definición 4.2 (Regularidad). Decimos que un punto factible es **regular**, si los gradientes de todas las restricciones *efectivas* en el punto son linealmente independientes (cuando los evaluamos en dicho punto).

Recordemos que podemos interpretar los términos que añadimos al lagrangiano como términos de *penalización*. Para poder escribir el lagrangiano siempre en la misma forma, en problemas con restricciones de desigualdad deberemos de escribir las desigualdades en forma *normalizada*.

Definición 4.3 (Problema de optimización normalizado). Decimos que un problema de optimización con restricciones de desigualdad está **normalizado** si todas las restricciones son de $\leq$ cuando el problema es de maximización, y de $\geq$ cuando es de minimización.

Definición 4.4 (Condiciones de Kuhn–Tucker para maximización). Consideremos el problema de maximización

$$\begin{array}{rrcl} \max\limits_{x\in\mathbb{R}^n} & f(x) & & \\ \text{s.a.} & g_1(x) & \leq & 0 \\ & g_2(x) & \leq & 0 \\ & \multicolumn{3}{l}{\dots\dots\dots\dots} \\ & g_m(x) & \leq & 0 \end{array}$$

Supongamos que las funciones f, g_1, g_2, ..., g_m, son diferenciables. Las **condiciones de Kuhn–Tucker** del problema son

$$\nabla f(x) = \lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) + \dots + \lambda_m \nabla g_m(x) \;\rightarrow\; \text{Tangencia}$$

$$\left.\begin{array}{lll} \lambda_1 \geq 0; & g_1(x) \leq 0; & \lambda_1 g_1(x) = 0 \\ \lambda_2 \geq 0; & g_2(x) \leq 0; & \lambda_2 g_2(x) = 0 \\ \dots\dots & \dots\dots & \dots\dots \\ \lambda_m \geq 0; & g_m(x) \leq 0; & \lambda_m g_m(x) = 0 \end{array}\right\} \rightarrow \begin{array}{l}\text{Factibilidad} \\ \text{y Exclusión}\end{array}$$

Definición 4.5 (Condiciones de Kuhn–Tucker para minimización). Consideremos el problema de minimización

$$\begin{array}{rrcl} \min\limits_{v\in\mathbb{R}^n} & f(v) & & \\ \text{s.a.} & g_1(v) & \geq & 0 \\ & g_2(v) & \geq & 0 \\ & \multicolumn{3}{l}{\dots\dots\dots\dots} \\ & g_m(v) & \geq & 0 \end{array}$$

Supongamos que las funciones f, g_1, g_2, ..., g_m, son diferenciables. Las **condiciones de Kuhn–Tucker** del problema son

$$\nabla f(v) = \lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) + \dots + \lambda_m \nabla g_m(x) \;\rightarrow\; \text{Tangencia}$$

$$\left.\begin{array}{lll} \lambda_1 \geq 0; & g_1(x) \geq 0; & \lambda_1 g_1(x) = 0 \\ \lambda_2 \geq 0; & g_2(x) \geq 0; & \lambda_2 g_2(x) = 0 \\ \dots\dots & \dots\dots & \dots\dots \\ \lambda_m \geq 0; & g_m(x) \geq 0; & \lambda_m g_m(x) = 0 \end{array}\right\} \rightarrow \begin{array}{l}\text{Factibilidad} \\ \text{y Exclusión}\end{array}$$

Al igual que en el caso con restricciones de igualdad, una manera conveniente de expresar las condiciones de Kuhn y Tucker es mediante el lagrangiano.

Definición 4.6 (Lagrangiano). El **lagrangiano**, o función de Lagrange, de un problema de optimización normalizado como los de las anteriores definiciones, es la función de $n+m$ argumentos:

$$L(x,\lambda) = f(x) - \lambda_1\, g_1(x) - \lambda_2\, g_2(x) - \cdots - \lambda_m\, g_m(x),$$

donde $x = (x_1, x_2, \ldots, x_n)$ y $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$.

Las condiciones de tangencia resultan de igualar a cero las derivadas parciales del lagrangiano con respecto a las variables de decisión.

Ejemplo 4.1. Consideremos el problema de minimización

$$\begin{aligned} \min_{(x,y)} \ & x\,y \\ \text{s.a.} \ & x^2 + y^2 \leq 2 \end{aligned}$$

Para hallar las condiciones de Kuhn y Tucker, en primer lugar lo escribiremos en forma normalizada:

$$\begin{aligned} \min_{(x,y)} \ & x\,y \\ \text{s.a.} \ & -x^2 - y^2 + 2 \geq 0 \end{aligned}$$

Por tanto: $f(x,y) = x\,y$ y $g(x,y) = -x^2 - y^2 + 2$. A continuación definimos el lagrangiano:

$$L(x,y,\lambda) = f(x,y) - \lambda\, g(x,y) = x\,y - \lambda\,\left(-x^2 - y^2 + 2\right)$$

Las condiciones de Kuhn y Tucker son:

$$\begin{aligned} \frac{\partial L}{\partial x} = y + 2\,\lambda\,x \ &= 0 && (a) \\ \frac{\partial L}{\partial y} = x + 2\,\lambda\,y \ &= 0 && (b) \\ \lambda \ &\geq 0 && (c) \\ x^2 + y^2 \ &\leq 2 && (d) \\ \lambda\,\left(-x^2 - y^2 + 2\right) \ &= 0 && (e) \end{aligned}$$

Para resolver estas condiciones, típicamente empezamos con las condiciones de tangencia. Igualando (a) y (b) obtenemos

$$(x-y)\,(1-2\,\lambda) = 0$$

Ahora hay dos posibilidades: o bien $x = y$, o bien $\lambda = 1/2$. Si exploramos las consecuencias de cada una de estas posibilidades, acabaremos por llegar a una solución, o bien a una contradicción. Comenzamos por la primera de ellas

$$x = y \xrightarrow{(a)} x\,(1 + 2\,\lambda) = 0 \xrightarrow{(c)} x = 0 = y \xrightarrow{(e)} \lambda = 0$$

Hemos llegado a la solución $(x,y,\lambda) = (0,0,0)$. Por otro lado:

$$\lambda = \frac{1}{2} \ \rightarrow \left\{ \begin{array}{l} \xrightarrow{(a)} \ x + y \quad = 0 \\ \xrightarrow{(e)} \ x^2 + y^2 = 2 \end{array} \right\} \rightarrow \left\{ \begin{array}{l} x \ = 1, \quad y = -1 \\ x \ = -1, \ y = 1 \end{array} \right.$$

Por tanto, los puntos $(1,-1,1/2)$ y $(-1,1,1/2)$ también solucionan las condiciones de Kuhn y Tucker. ⊛

Las condiciones de exclusión implican que o bien una restricción es efectiva o bien el multiplicador correspondiente es cero. En general, sólo una de las dos cosas es cierta, aunque no tiene por qué ser así: es fácil encontrar ejemplos en que ambas cosas pueden suceder simultáneamente. De hecho, estos casos pueden dar problemas, como cuando deseamos hacer un análisis de sensibilidad. Llamaremos **soluciones degeneradas** aquellas en que se da esta circunstancia.

4.2 Condiciones de primer orden

Teorema 4.1 (Condiciones de Primer Orden). Consideremos un problema de optimización con restricciones de desigualdad en que todas las funciones son diferenciables con continuidad. Entonces todo maximizador o minimizador local que sea un punto regular satisface las condiciones de Kuhn-Tucker.

Observación: El teorema *no* afirma que, si las condiciones de Kuhn y Tucker tienen soluciones, *todas* ellas solucionarán el problema de optimización, ni tan solo que dicho problema tenga solución alguna. Sólo afirma que las condiciones de Kuhn y Tucker detectarán cualquier maximizador o minimizador local.

4.3 Condiciones de segundo orden

Teorema 4.2 (Condiciones Necesarias de Segundo Orden). Supongamos que las funciones f, g_1, $g_2, \ldots, g_m$ son diferenciables dos veces con continuidad. Entonces tenemos:

(i) Supongamos que x^* es un maximizador local y un punto regular de las restricciones. Entonces existe λ^* tal que (x^*, λ^*) satisfacen las condiciones de Kuhn y Tucker, y además la matriz hessiana $D_x^2 L(x^*, \lambda^*)$ es semidefinida negativa sobre el plano tangente a las restricciones efectivas.

(ii) Supongamos que x^* es un minimizador local y un punto regular de las restricciones. Entonces existe λ^* tal que (x^*, λ^*) satisfacen las condiciones de Kuhn y Tucker, y además la matriz hessiana $D_x^2 L(x^*, \lambda^*)$ es semidefinida positiva sobre el plano tangente a las restricciones efectivas.

Teorema 4.3 (Condiciones Suficientes de Segundo Orden). Supongamos que las funciones f, g_1, $g_2, \ldots, g_m$ son diferenciables dos veces con continuidad, y que existe λ^* tal que (x^*, λ^*) satisfacen las condiciones de Kuhn y Tucker. Entonces tenemos:

(i) Si la matriz hessiana $D_x^2 L(x^*, \lambda^*)$ es definida negativa sobre el plano tangente a todas las restricciones i tales que $\lambda_i^* > 0$, entonces x^* es un maximizador local estricto.

(ii) Si la matriz hessiana $D_x^2 L(x^*, \lambda^*)$ es definida positiva sobre el plano tangente a todas las restricciones i tales que $\lambda_i^* > 0$, entonces x^* es un minimizador local estricto.

Notemos que, para las condiciones suficientes, hemos de considerar no las restricciones efectivas, sino aquellas que tienen asociado un multiplicador estrictamente positivo, es decir, las restricciones efectivas no degeneradas.

4.4 Criterios de globalidad. Cuasiconcavidad y cuasiconvexidad

Teorema 4.4 (Globalidad con funciones cóncavas y convexas). En un problema normalizado de optimización con restricciones de desigualdad en que todas las funciones son diferenciables, toda solución de las condiciones de Kuhn-Tucker es una solución global siempre que se den las siguientes condiciones:

(i) Si el problema es de maximización, la función objetivo es cóncava y las restricciones son funciones convexas.

(ii) Si el problema es de minimización, la función objetivo es convexa y las restricciones son funciones cóncavas.

DEMOSTRACIÓN: Consideraremos solamente el caso de maximización. Supongamos que (x^*, λ^*) satisfacen las condiciones de Kuhn y Tucker. Un par de observaciones elementales: (1) el negativo de una función convexa es una función cóncava; y (2) la suma de funciones cóncavas es una función cóncava. En consecuencia, la concavidad de f y la convexidad de las restricciones implica que el lagrangiano $L(x, \lambda^*) = f(x) - \sum_{i=1}^{m} \lambda_i^* \, g_i(x)$ es una función cóncava en la variable x, ya que $\lambda^* \geq 0$. Por tanto, dado cualquier punto factible $\bar{x}$ tenemos

$$L(\bar{x}, \lambda^*) - L(x^*, \lambda^*) \leq \nabla_x L(x^*, \lambda^*) \, (\bar{x} - x^*)$$

Pero la condición de tangencia implica que $\nabla_x L(x^*, \lambda^*) = 0$, por lo cual

$$L(\bar{x}, \lambda^*) \leq L(x^*, \lambda^*)$$

Substituyendo en las definiciones de los lagrangianos respectivos

$$f(\bar{x}) - \sum_{i=1}^{m} \lambda_i^* \, g_i(\bar{x}) \leq f(x^*) - \sum_{i=1}^{m} \lambda_i^* \, g_i(x^*)$$

Pero la condición de exclusión implica que, para cada i, $\lambda_i^* \, g_i(x^*) = 0$, y por tanto

$$f(\bar{x}) - \sum_{i=1}^{m} \lambda_i^* \, g_i(\bar{x}) \leq f(x^*)$$

También sabemos que, para cada i, $\lambda_i^* \geq 0$ y $g_i(\bar{x}) \leq 0$, o sea que $\lambda_i^* \, g_i(\bar{x}) \leq 0$ (recordemos que $\bar{x}$ no tiene por qué satisfer la condición de KT, y por tanto este último producto no tiene por qué ser cero), o sea que

$$f(\bar{x}) \leq f(\bar{x}) - \sum_{i=1}^{m} \lambda_i^* \, g_i(\bar{x}) \leq f(x^*)$$

Lo que concluye la demostración. □

Para extender el resultado que acabamos de ver, debemos acudir a generalizaciones de los conceptos de concavidad y convexidad.

Definición 4.7 (Cuasiconcavidad y cuasiconvexidad). Sea X un subconjunto convexo de $\mathbb{R}^n$.

(i) Decimos que la función $f : X \to \mathbb{R}$ es **cuasicóncava** si, para cada $x \in X$, el contorno superior $\{y \in X : f(y) \geq f(x)\}$ es un conjunto convexo.

(ii) Decimos que la función $f : X \to \mathbb{R}$ es **cuasiconvexa** si, para cada $x \in X$, el contorno inferior $\{y \in X : f(y) \leq f(x)\}$ es un conjunto convexo.

Cuando f es una función continua la definición anterior puede ser hecha indistintamente en términos de los contornos definidos con desigualdades débiles o estrictas. La comprobación de la siguiente proposición consiste en un ejercicio sencillo.

Proposición 4.1 (Concavidad y cuasiconcavidad). Toda función cóncava es cuasicóncava, y tota función convexa es cuasiconvexa.

Las siguiente equivalencias son básicamente reescrituras de las definiciones respectivas.

Proposición 4.2 (Definición algebraica de cuasiconcavidad (cuasiconvexidad)). Sea X un subconjunto convexo de $\mathbb{R}^n$.

(i) La función $f : X \to \mathbb{R}$ es cuasicóncava si, y sólo si, para cada x e y de X, y para cada $0 \leq t \leq 1$, se cumple que $f[tx + (1-t)y] \geq \min\{f(x), f(y)\}$.

(ii) La función $f : X \to \mathbb{R}$ es cuasiconvexa si, y sólo si, para cada x e y de X, y para cada $0 \leq t \leq 1$, se cumple que $f[tx + (1-t)y] \leq \max\{f(x), f(y)\}$.

Recordemos que, en un problema de optimización, una transformación estrictamente creciente de la función objetivo no altera la solución (maximizador o minimizador), puesto que no cambia el orden relativo de los valores.

Proposición 4.3 (Transformaciones crecientes y cuasiconcavidad). (i) Toda transformación estrictamente creciente de una función cuasicóncava resulta en una nueva función cuasicóncava. En particular, una transformación estrictamente creciente de una función cóncava es una función cuasicóncava (no necesariamente cóncava).

(ii) Toda transformación estrictamente creciente de una función cuasiconvexa resulta en una nueva función cuasiconvexa. En particular, una transformación estrictamente creciente de una función convexa es una función cuasiconvexa (no necesariamente convexa).

Restricciones de desigualdad $\leq$ dadas por funciones cuasiconvexas, dan lugar a un conjunto factible convexo. Análogamente, restricciones de desigualdad $\geq$ dadas por funciones cuasicóncavas, dan lugar a un conjunto factible convexo.

Proposición 4.4 (Función objetivo cóncava y restricciones cuasiconvexas). Supongamos que en un problema normalizado de optimización con restricciones de desigualdad todas las funciones son diferenciables con continuidad, y se da uno de los dos siguientes casos:

(i) El problema es de maximización, la función objetivo es cóncava y las restricciones son funciones cuasiconvexas.

(ii) El problema es de minimización, la función objetivo es convexa y las restricciones son funciones cuasicóncavas.

Entonces toda solución de las condiciones de Kuhn-Tucker es una solución global del problema.

Estos resultados se generalizan a funciones objetivo cuasicóncavas o cuasiconvexas, con una excepción notable.

Proposición 4.5 (Función objetivo cuasicóncava y restricciones cuasiconvexas). Supongamos que en un problema normalizado de optimización con restricciones de desigualdad todas las funciones son diferenciables con continuidad, y se da uno de los dos siguientes casos:

(i) El problema es de maximización, la función objetivo es cuasicóncava y las restricciones son funciones cuasiconvexas.

(ii) El problema es de minimización, la función objetivo es cuasiconvexa y las restricciones son funciones cuasicóncavas.

Entonces toda solución de las condiciones de Kuhn-Tucker es una solución global del problema, siempre que no tenga todas las derivadas parciales iguales a cero.

Ejemplo 4.2. Consideremos el problema

$$\begin{aligned} \max_{(x,y)} \ & x\, y \\ \text{s.a.} \ & x^2 + y^2 \leq 2 \\ & x \geq 0 \\ & y \geq 0 \end{aligned}$$

Normalizando,

$$\begin{aligned} \max_{(x,y)} \ & x\, y \\ \text{s.a.} \ & x^2 + y^2 - 2 \leq 0 \\ & -x \leq 0 \\ & -y \leq 0 \end{aligned}$$

La función objetivo es cuasicóncava y las restricciones normalizadas son todas funciones convexas. Las condiciones de Kuhn-Tucker tienen como solución los dos puntos donde (x, y, λ) valen $(1, 1, 1/2)$ y $(0, 0, 0$. Tenemos que $\nabla f(1, 1) = (1, 1) \neq (0, 0)$, pero $\nabla f(0, 0) = (0, 0)$. El teorema anterior implica que el primer punto es un maximizador global, pero no se pronuncia sobre el segundo. De hecho, $(0, 0)$ *no* es un maximizador, sino que se trata de un punto de ensilladura de f, que aparece aquí puesto que todo punto crítico (donde las derivadas parciales se anulan) de una función es siempre solución de las condiciones de KT, si está dentro del conjunto factible. En una función cóncava esto no es problema, ya que todo punto donde todas las derivadas parciales son zero es automáticamente un maximizador global, pero esto no es necesariamente cierto cuando una función es cuasicóncava, como el presente ejemplo ilustra. ⊛

4.5 Problemas con restricciones de igualdad y de desigualdad

Cuando hay restricciones tanto de igualdad como de desigualdad, las condiciones de Khun-Tucker son las mismas, con la excepción que los multiplicadores asociados a restricciones de igualdad no tienen por qué ser negativos. Con esta modificación, todos los resultados *locales* que hemos mostrado son todavía válidos para este caso.

Por lo que respecta a las condiciones de globalidad en términos de concavidad y convexidad, los teoremas anteriores siguen siendo aplicables siempre que las restricciones de igualdad sean lineales (podem ver esto escribiendo una igualdad como dos desigualdades con signos opuestos: $a = b \iff a \leq b$ y $a \geq b$). Si hay restricciones de igualdad que no son lineales, podemos aplicar un argumento de globalidad como el usado en el caso en que solamente había restricciones de igualdad: consideramos para cada solución de las condiciones de KT el lagrangiano después de substituir los valores óptimos de los multiplicadores, y comprobamos si la función resultante es cóncava (para maximización) o convexa (para minimización); si la respuesta es afirmativa, entonces la solución correspondiente es global.

5 Análisis de sensibilidad

5.1 Problemas paramétricos de optimización. El teorema del máximo

En los problemas de decisión individual, hay una serie de variables que están bajo control del decisor, las *variables de decisión*, y otras que vienen dadas externamente y que definen el entorno del decisor, es decir, su conjunto factible y su función objetivo. Llamamos *parámetros* aquellas variables que no están sometidas al control del decisor. Por ejemplo, si una empresa toma decisiones en relación a un cierto mercado, su propio precio es una variable de decisión, y los precios de las demás empresas son parámetros. En los problemas paramétricos de optimización, tan importante como conocer la solución óptima es saber como cambiará ésta frente a variaciones en los parámetros.

Definición 5.1 (Variable de decisión). En un problema de optimización, las **variables de decisión** son aquellas respecto a las cuales estamos optimizando.

Definición 5.2 (Parámetro). En un problema de optimización, un **parámetro** es toda variable que se toma como dada. Es decir, el parámetro es una variable cuyo valor se supone fijo en el momento de hallar el óptimo.

Definición 5.3 (Funciones de decisión). Llamamos **funciones de decisión** aquellas funciones que expresan el valor óptimo de las variables de decisión en términos de los parámetros.

Definición 5.4 (Función de valor). Llamamos **función de valor** aquella función que expresa el valor óptimo de la función objetivo en términos de los parámetros.

Notemos que un problema de optimización que tenga solución para todo valor de los parámetros tendrá siempre una función de valor bien definida, puesto que el valor máximo o mínimo es único, pero no necesariamente tendrá funciones de decisión bien definidas, puesto que un máximo o mínimo puede corresponder a distintos maximizadores o minimizadores. En general, lo que siempre estará bien definido serán las **correspondencias de decisión**, que contienen todos los maximizadores o minimizadores que corresponden a cada valor de los parámetros. Las funciones de decisión están bien definidas cuando la correspondencias de decisión tienen un único elemento para cada valor de los parámetros.

Un resultado básico en problemas paramétricos de optimización es el **teorema del máximo**, que da condiciones bajo las cuales la función de valor y las funciones de decisión (en caso de estar bien definidas) son continuas.

Supongamos que las funciones numéricas (i.e. con valores reales) f, g_i (para $1 \leq i \leq m$), y h_j (para $1 \leq j \leq r$), están definidas sobre un dominio $X \times P$, donde $x \in X \subseteq \mathbb{R}^n$ son las variables de decisión, y $p \in P \subseteq \mathbb{R}^k$ son los parámetros. Supongamos que tenemos planteado el problema de optimizar (maximizar o minimizar) la función f sujeta a restricciones de desigualdad ($\leq$ o $\geq$) para todas las funciones g_i ($1 \leq i \leq m$) y restricciones de igualdad para las funciones h_j ($1 \leq j \leq r$). Sea $v(p)$ la función de valor del problema.

Teorema 5.1 (Teorema del máximo). Supongamos que el conjunto X es compacto (cerrado y acotado), y que las funciones f, g_i ($1 \leq i \leq m$), y h_j ($1 \leq j \leq r$) son continuas con respecto a sus argumentos (x, p). Supongamos adicionalmente que, por cada valor de $p \in P$, existe por lo menos una $x \in X$ que satisface todas las restricciones. Entonces la función de valor $v(p)$ está bien definida y es continua en p. Adicionalmente, si para cada $p \in P$ hay un único maximizador o minimizador $x(p)$ (la correspondencia de decisión es una función), entonces $x(p)$ es también una función continua.

5.2 Análisis de sensibilidad: el teorema de la función implícita

En un problema paramétrico de optimización, el *análisis de sensibilidad* consiste en el estudio del comportamiento de las funciones de decisión y la función de valor frente a cambios en el valor de los parámetros. En general, no hay forma de expresar estas funciones en forma explícita, y por tanto deberemos analizar su comportamiento a partir de las condiciones de primer orden, que expresan en forma implícita la relación entre las variables de decisión y los parámetros.

Notemos que, en las condiciones de primer orden de un problema de optimización con restricciones, *los multiplicadores de Lagrange son variables endógenas*, que dependen de los parámetros en la misma forma en que lo hacen las variables de decisión.

La herramienta matemática que justifica el análisis de estas funciones definidas implícitamente es el **teorema de la función implícita**.

Consideremos una función $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. Dados $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ y $p = (p_1, p_2, \ldots, p_m) \in \mathbb{R}^m$, escribamos $f(x, p) = 0$ para referirnos al sistema de ecuaciones:

$$
\begin{aligned}
f_1(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) &= 0 \\
f_2(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) &= 0 \\
\cdots\ \cdots\ \cdots& \\
f_n(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) &= 0
\end{aligned}
$$

Notemos que las x son las variables endógenas y las p los parámetros o variables exógenas, y el hecho de tener tantas ecuaciones como variables endógenas nos permite plantearnos la posibilidad de expresar las x en función de los parámetros p.

Designaremos mediante $D_x f(x, p)$ la matriz (cuadrada, de dimensión $n \times n$) de derivadas parciales de las funciones f_k con respecto a las variables x_i:

$$
D_x f(x, p) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}
$$

y con $D_p f(x, p)$ la matriz de derivadas parciales de las funciones f_k con respecto a las variables p_j:

$$
D_p f(x, p) = \begin{pmatrix} \frac{\partial f_1}{\partial p_1} & \frac{\partial f_1}{\partial p_2} & \cdots & \frac{\partial f_1}{\partial p_m} \\ \frac{\partial f_2}{\partial p_1} & \frac{\partial f_2}{\partial p_2} & \cdots & \frac{\partial f_2}{\partial p_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial p_1} & \frac{\partial f_n}{\partial p_2} & \cdots & \frac{\partial f_n}{\partial p_m} \end{pmatrix}
$$

Naturalmente, $D_x f(x, p)$ y $D_p f(x, p)$ son submatrices de la matriz jacobiana $Df(x, p)$.

Teorema 5.2 (Teorema de la función implícita). Sea $A \subset \mathbb{R}^n \times \mathbb{R}^m$ un conjunto abierto, y sea $f : A \to \mathbb{R}^n$ una función con derivadas parciales de orden $t \geq 1$ que son funciones continuas. Supongamos que hay un cierto punto $(\bar{x}, \bar{p}) \in A$ tal que $f(\bar{x}, \bar{p}) = 0$. Supongamos que la matriz de derivadas parciales $D_x f(\bar{x}, \bar{p})$ tiene un determinante distinto de cero (i.e., es invertible). Entonces hay un entorno abierto $U \subset \mathbb{R}^n$ de $\bar{x}$, un entorno abierto $V \subset \mathbb{R}^m$ de $\bar{p}$, y una (única) función $g : V \to U$, tales que $\bar{x} = g(\bar{p})$ y

$$
f\big(g(p), p\big) = 0
$$

para todo $p \in V$. Además, la función g tiene derivadas parciales de orden t, y éstas son funciones continuas.

El teorema garantiza que, localmente alrededor del punto $(\bar{x}, \bar{p})$, podemos expresar $x = g(p)$, donde la función g tiene el mismo grado de diferenciabilidad que la f. A partir de la existencia de esta función, podemos diferenciar con respecto a p la identidad:

$$
f\big(g(p), p\big) = 0
$$

y obtenemos:

$$D_x f\big(g(p), p\big)\, Dg(p) + D_p f\big(g(p), p\big) = 0$$

De donde podemos resolver para las derivadas de la función g:

$$Dg(p) = -\big(D_x f\big(g(p), p\big)\big)^{-1} D_p f\big(g(p), p\big)$$

Aplicando el teorema de la función implícita a las condiciones de primer orden, podemos analizar la variación de las funciones de decisión. Análogamente, en modelos económicos en que el equilibrio viene expresado por una serie de ecuaciones que relacionan las variables endógenas con las variables exógenas, el **análisis de estática comparativa** consiste en analizar el efecto sobre las variables endógenas de cambios en el valor de las variables exógenas: el teorema de la función implícita justifica la realización a nivel local de este tipo de análisis mediante la diferenciación.

5.3 La función de valor y el teorema de la envolvente

La función de valor puede depender de los parámetros directamente y depende también de las variables de decisión, que a su vez dependen de los parámetros. Para analizar la variación (derivadas) de esta función en relación a los parámetros, hay que tener en cuenta no sólo la influencia directa de los mismos, sino también su influencia indirecta a través de las funciones de decisión. La técnica matemática que nos permite recoger ambas influencias es la **regla de la cadena**.

Teorema 5.3 (Regla de la Cadena). Sean $A \subset \mathbb{R}^n$ y $B \subset \mathbb{R}^m$ conjuntos abiertos, y sea $f : A \to \mathbb{R}^m$ una función diferenciable tal que $f(A) \subset B$. Entonces, si $g : B \to \mathbb{R}^p$ es una función diferenciable, la función compuesta $g \circ f : A \to \mathbb{R}^p$ es también diferenciable, y se cumple que

$$D(g \circ f)(x) = Dg[f(x)] \circ Df(x)$$

Este teorema afirma que *la derivada de la función compuesta es igual a la composición de derivadas*. Teniendo en cuenta que las derivadas son funciones lineales representadas por matrices, esta composición de derivadas se traduce en un producto de matrices.

Ejemplo 5.1. Supongamos que tenemos una función compuesta de la forma $g(x, y) = h[f(x, y)]$. La regla de la cadena dice que sus derivadas parciales son:

$$\frac{\partial g}{\partial x}(x, y) = h'[f(x, y)]\, \frac{\partial f}{\partial x}(x, y)$$
$$\frac{\partial g}{\partial y}(x, y) = h'[f(x, y)]\, \frac{\partial f}{\partial y}(x, y)$$

Esquemáticamente, si $z = f(x, y)$, podemos escribir

$$\frac{\partial g}{\partial x} = h'(z)\, \frac{\partial f}{\partial x}$$
$$\frac{\partial g}{\partial y} = h'(z)\, \frac{\partial f}{\partial y}$$

⊛

Ejemplo 5.2. Supongamos que tenemos una función compuesta de la forma $v(p) = f(x, y) = f[g(p), h(p)]$. La regla de la cadena afirma que su derivada es

$$v'(p) = \frac{\partial f}{\partial x} g'(p) + \frac{\partial f}{\partial y} h'(p)$$

La intuición de este resultado es la siguiente: f es función tanto de x como de y, las cuales a su vez dependen de p. La variación total de f con respecto a p (en ocasiones denominada la *derivada total*), viene dada por la suma de la influencia de p a través de cada una de las variables x e y. ⊛

Sea $f(x, p)$ una función diferenciable de las variables reales x y p. Dado un cierto valor de p, digamos $\bar{p}$, consideremos el problema

$$\max_x \quad f(x, \bar{p})$$

Supongamos que la solución a este problema viene descrita por la condición de primer orden:

$$\frac{\partial f}{\partial x}(x, \bar{p}) = 0 \tag{1}$$

Notemos que esta condición de primer orden no es más que una ecuación que define implícitamente la función $x(p)$. Supongamos que nos interesa analizar la función de valor $v(p) = f[x(p), p]$. Diferenciando la función de valor, obtenemos, por aplicación de la regla de la cadena:

$$\frac{dv}{dp} = \frac{\partial f}{\partial x}\frac{dx}{dp} + \frac{\partial f}{\partial p} \tag{2}$$

Pero la condición de primer orden que hemos escrito en (1) implica que el primer término de la suma es cero, por lo cual queda

$$\frac{dv}{dp} = \frac{\partial f}{\partial p} \tag{3}$$

Este resultado es conocido como *teorema de la envolvente*. Notemos que el parámetro p influye en la función de valor directamente, en tanto que argumento de f, y también indirectamente, a través de su influencia sobre la función de decisión $x(p)$. Cuando hemos aplicado la regla de la cadena a f, en la ecuación (2), nos ha quedado un término debido al efecto indirecto y otro debido al efecto directo. Entonces hemos observado que las condiciones de primer orden implican que el efecto indirecto es nulo, y por tanto la derivada de la función de valor depende solamente del efecto directo.

La figura 5.3 ilustra por qué este resultado es cierto. Dado un valor fijo del parámetro, p_1, sea x_1 el valor óptimo de la variable de decisión. Análogamente, sea x_2 el valor óptimo que corresponde a un cierto valor p_2. En el gráfico, hemos puesto los valores de p en el eje de las abcisas. Entonces hemos dibujado la función de valor $v(p)$, y también las funciones $h_1(p) = f(x_1, p)$ y $h_2(p) = f(x_2, p)$, que no son más que la función f cuando fijamos los valores de x en x_1 y x_2, respectivamente.

Como x_1 es óptima cuando el parámetro vale p_1, tenemos que $v(p_1) = f(x_1, p_1) = h_1(p_1)$. Para otros valores de p, x_1 ya no tiene per qué ser óptima, y por tanto $h_1(p) = f(x_1, p) \leq v(p)$. Es decir, la gráfica de h_1 toca a la de v en el punto p_1, y está por debajo en los demás puntos. Como las dos gráficas son tangentes en el punto p_1, sus derivadas en este punto coinciden, es decir:

$$v'(p_1) = h_1'(p_1) = \frac{\partial f}{\partial p}(x_1, p_1)$$

El teorema de la envolvente consiste en esta igualdad de las derivadas.

El teorema de la envolvente afirma que, cuando diferenciamos la función de valor, podemos olvidarnos de los efectos indirectos de los parámetros sobre la función objetivo a través de las funciones de decisión,

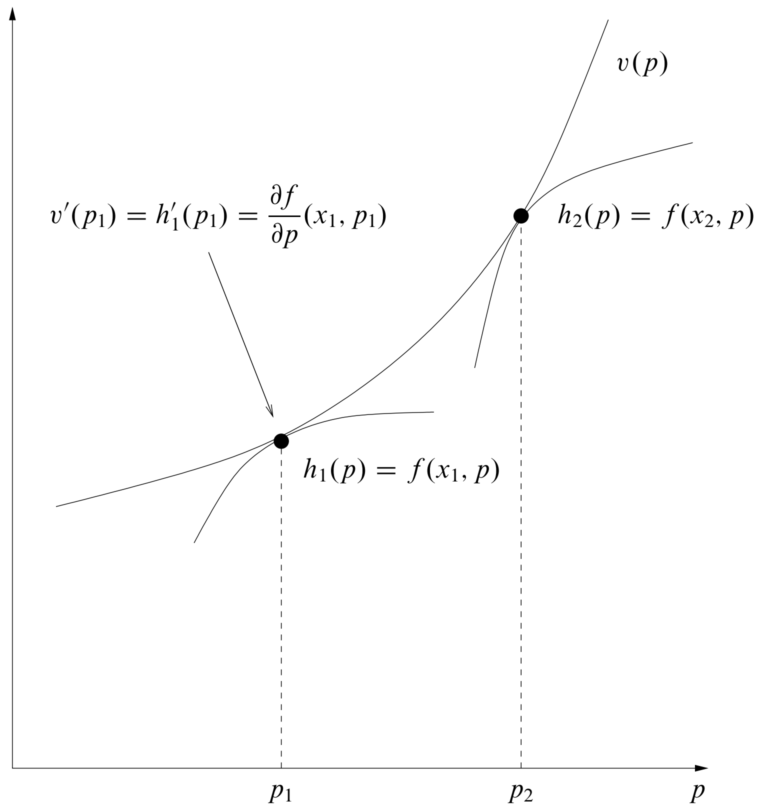

Figure 1: Ilustración del teorema de la envolvente.

ya que las condiciones de primer orden implican que estos efectos son nulos, y sólo quedan los efectos directos. Un hecho notable es que este resultado se extiende también a problemas con restricciones, si en lugar de la función f consideramos los efectos directos de los parámetros *sobre el lagrangiano*. De hecho, la interpretación de los multiplicadores de Lagrange como *precios sombra* en determinados modelos económicos, se puede obtener como una consecuencia inmediata del teorema de la envolvente. Pasemos a una exposición formal de las distintas variantes del teorema.

Diremos que una función que toma valores reales es de clase C^m si sus derivadas parciales de orden m existen y son funciones continuas.

Sea f una función definida sobre un dominio (conjunto abierto y conexo) contenido en $\mathbb{R}^n \times \mathbb{R}^k$, y que toma valores reales. Supongamos que f es de clase C^2 en relación al primer argumento y $D_x f(x, p)$ es de clase C^1 en relación al segundo argumento. Consideremos el problema de maximizar o minimizar la función f en relación al primer argumento.

Teorema 5.4 (Problemas sin restricciones). Sea $x^* \in \mathbb{R}^n$ una solución local del problema de optimización, dado un cierto valor de $p^* \in \mathbb{R}^k$. Supongamos que se satisface la condición suficiente de segundo orden para un óptimo local estricto. Entonces existe un entorno de p^* y una función diferenciable $x(p)$ definida en dicho entorno, tal que $x(p)$ es un óptimo local estricto de f dado p. Si definimos $v(p) = f[x(p), p]$, entonces se cumple la condición de envolvente:

$$Dv(p) = D_p f(x(p), p)$$

Sea f una función con las características descritas más arriba, y sean g_i, para $i \in \{1, 2, \ldots, m\}$, funciones reales definidas sobre el mismo dominio y satisfaciendo las mismas condiciones de diferenciabilidad. Consideremos el problema de buscar extremos locales de f en relación al primer argumento, con la restricción de que cada g_i es igual a 0.

Teorema 5.5 (Problemas con restricciones de igualdad). Sea (x^*, λ^*) una solución local del problema de optimización, dado un cierto valor de $p^* \in \mathbb{R}^k$, tal que x^* es un punto regular de las restricciones. Supongamos que se satisface la condición suficiente de segundo orden para un óptimo local estricto. Entonces existe un entorno de p^* y existen funciones diferenciables $x(p)$ y $\lambda(p)$, definidas en dicho entorno, tales que $x(p)$ es un óptimo local estricto de f sobre el conjunto definido per las restricciones, dado p. Si definimos $v(p) = f[x(p), p]$, tenemos que se cumple la condición de envolvente:

$$Dv(p) = D_p L(x(p), \lambda(p), p)$$

Las condiciones de desigualdad las enunciaremos para el caso de maximización. El caso de minimización es similar. Supongamos que las funciones son todas tal como las hemos descrito más arriba, pero consideremos ahora el problema de maximizar f sujeta a las restricciones de que cada g_i es menor o igual a 0.

Teorema 5.6 (Problemas con restricciones de desigualdad). Sea (x^*, λ^*) una solución local del problema de maximización, dado un cierto valor de $p^* \in \mathbb{R}^k$, tal que x^* es un punto regular de las restricciones. Supongamos que se satisface la condición suficiente de segundo orden para un máximo local estricto. Supongamos, adicionalmente, que los multiplicadores asociados a las restricciones efectivas son todos distintos de zero (condición de no degeneración). Entonces existe un entorno de p^* y existen funciones diferenciables $x(p)$ y $\lambda(p)$, definidas en dicho entorno, tales que $x(p)$ es un óptimo local estricto de f sobre el conjunto definido por las restricciones, dado p. Si definimos $v(p) = f[x(p), p]$, tenemos que se cumple la condición de envolvente:

$$Dv(p) = D_p L(x(p), \lambda(p), p)$$

Si hay restricciones tanto de igualdad como de desigualdad, la condición de no degeneración se aplica solamente a las restricciones de desigualdad.

5.3.1 Limitaciones del teorema de la envolvente

En el caso de restricciones de desigualdad, hemos requerido que la solución sea no degenerada. La solución es degenerada cuando el multiplicador asociado a una restricción efectiva es cero. En este caso, es posible que los multiplicadores ni tan sólo sean funciones continuas de los parámetros, y por tanto no es válido diferenciar las condiciones de primer orden ni podemos aplicar el teorema de la envolvente. Ilustremos esto mediante un ejemplo.

Ejemplo 5.3. Consideremos el siguiente problema de maximización

$$\begin{aligned} \max_{(x,y)} \quad & -(x-1)^2 + y \\ \text{s.a.} \quad & -x + y \leq 0 \\ & x + y \leq 2 \\ & y \leq p \end{aligned}$$

Alrededor del valor del parámetro $p = 1$. Sean λ, μ y γ los respectivos multiplicadores asociados a las restricciones.

Cuando $p = 1$, hay dos soluciones de las condiciones de Kuhn–Tucker:

$$x = 1 \quad y = 1 \quad \lambda = 0 \quad \mu = 0 \quad \gamma = 1 \tag{4}$$

$$x = 1 \quad y = 1 \quad \lambda = \frac{1}{2} \quad \mu = \frac{1}{2} \quad \gamma = 0 \tag{5}$$

Notemos que tanto (4) como (5) se refieren a la misma solución en términos de las variables de decisión, y que ambas soluciones son degeneradas. Una representación gráfica debería ayudar al lector.

Notemos también que, si p crece, nos interesaría tomar la solución (5), pero en cambio si p decrece la que nos interesa es (4); lo cual significa que los multiplicadores son discontinuos, sea cual sea la solución que tomemos. En particular, $\frac{\partial L}{\partial p}$ es γ, y según p aumente o disminuya nos interesaría tomar uno u otro de los valores que γ toma en las dos soluciones. En definitiva, ni el teorema de la envolvente es aplicable, ni podemos diferenciar las condiciones de primer orden. ⊛