Optimización

Ricard Torres Bargalló

Ejercicios a cargo de
Margarida Corominas Bosch
Anna Espinal Berenguer

PID_00186450

Universitat Oberta de Catalunya

Índice

Introducción

En la parte de Matemáticas I dedicada al cálculo univariante, nos encontramos con los conceptos básicos de la teoría de la optimización. Aquí haremos un tratamiento más sistemático del asunto. La teoría de la optimización se ocupa de todo lo que guarda relación con la búsqueda de valores máximos y mínimos de funciones. Nosotros clasificaremos aquí la optimización en tres grandes apartados:

1. Optimización sin restricciones.

2. Con restricciones de igualdad.

3. Con restricciones de desigualdad.

El criterio de clasificación depende de las **restricciones** que tiene un determinado problema, que son condiciones, en forma de igualdad o de desigualdad, que deben cumplir las variables. Establecemos esta clasificación porque las técnicas de resolución que usaremos cambian según el tipo de problema. Nuestro enfoque más bien trata de resaltar la lógica común que consta en los diferentes métodos empleados.

Objetivos

A lo largo de este módulo podréis alcanzar los objetivos siguientes:

1. Aprender a encontrar analíticamente la solución de problemas de optimización, tanto con restricciones como sin ellas.

2. Conocer la regla de los multiplicadores de Lagrange para resolver problemas con restricciones de igualdad, y saber cómo se formulan las condiciones de Kuhn y Tucker, que describen las soluciones de problemas en los que las restricciones son de desigualdad.

3. Conocer las condiciones de segundo orden para determinar el carácter local de los puntos críticos, pero sobre todo las condiciones de concavidad y/o convexidad que aseguran el carácter global de las soluciones.

4. Saber justificar de forma intuitiva y gráfica los métodos empleados, con el objetivo de que, ante un problema que no se amolde exactamente a la teoría que habéis aprendido, seáis capaces de tener criterio para razonar por qué un cierto método puede ser válido para encontrar la solución, o por qué no.

1. Conceptos básicos

En muchas circunstancias de nuestra vida nos encontramos con qué debemos seleccionar, de entre un número dado de alternativas, aquella que más nos interesa. Supongamos que ante un problema de este tipo, podemos asociar a cada una de las alternativas posibles un índice numérico que expresa su deseabilidad relativa, de tal modo que, para saber cuál de las dos alternativas es mejor, sólo tenemos que comparar la magnitud de sus índices respectivos. Nos encontramos, en este caso, con un problema de optimización. El problema es de maximización si valores mayores que el índice indican alternativas mejores, y es de minimización cuando lo que preferimos son valores menores que el índice.

Ejemplo 1.1. Supongamos que hacemos corresponder 0 a un suspenso, 1 a un aprobado justo, 2 a un aprobado, 3 a un notable y 4 a un excelente. Si queremos encontrar al estudiante que tiene el mejor expediente de la UOC, sólo tenemos que asociar a cada estudiante la media de sus notas y resolver el correspondiente problema de maximización.

Por cierto...

... esta asociación de números a notas no es arbitraria. Si alguno de vosotros se anima, cuando acabe los estudios, a hacer un máster o un doctorado en un país como por ejemplo Estados Unidos, verá que en la solicitud de admisión al programa le piden un índice de notas como éste.

Ejemplo 1.2. Pepito es un gran aficionado a los coches potentes, pero es más bien tacaño. Se está planteando comprar un coche, y por ello buscará aquél para el que sea más bajo el cociente entre el precio de compra y la potencia (medida en CV). Lo que hará será resolver un problema de minimización.

Ejemplo 1.3. Los servicios de meteorología registran cada año la pluviosidad total que ha habido en toda una serie de observatorios que tienen repartidos por el territorio. Encontrar el lugar donde ha llovido más es un problema de maximización, y el lugar donde ha llovido menos, uno de minimización. Lo mismo se aplica al registro de otras magnitudes, como por ejemplo las temperaturas.

Ejemplo 1.4. De entre todos los rectángulos que tienen un perímetro dado (digamos 1), queremos encontrar aquel que tiene el área máxima. Esto es un problema de maximización. Más adelante veremos cómo podemos resolverlo. Observamos, de momento, que hay una diferencia esencial entre este ejemplo y los anteriores que hemos visto: en los otros debíamos seleccionar entre un número finito de alternativas; por lo tanto, su resolución no requería ninguna técnica más sofisticada que la de irlas comparando, y armarse de paciencia si había muchas; en este ejemplo, en cambio, esta técnica no nos llevaría demasiado lejos, ya que hay un número infinito de alternativas (la longitud de la base puede ser cualquier número real entre 0 y1/2).

Ejemplo 1.5. María junta caracolas en la playa y después las vende a los turistas que pasean por allí. Con el tiempo, se ha dado cuenta de que la cantidad de caracolas que vende depende del precio que pide por las mismas, de acuerdo con la siguiente relación: $q = 100 - 2p$, en la que q representa la cantidad y p el precio. María querría saber cuál es el precio que debería poner para tener un ingreso más elevado. Aquí el índice que nos permite comparar alternativas es el ingreso, es decir, el precio multiplicado por la cantidad:

$$\text{Ingreso} = I(p) = pq = p(100 - 2p) = 100p - 2p^2$$

Si consideramos que el precio puede ser cualquier número real positivo, nos encontramos de nuevo con un problema en el que se encuentra un número infinito de alternativas.

En la base de muchos modelos económicos que tratan de hacer predicciones sobre lo que pasará, se encuentra el supuesto de que los agentes elegirán, de todas aquellas alternativas que tienen a su alcance, aquella o aquellas que les den mejores resultados. Dicho en otras palabras, estos modelos económicos suponen que los agentes resuelven algún tipo de problema de optimización. Pero la optimización está presente no sólo en modelos abstractos, sino en aspectos tan concretos como en la planificación de la producción, la política de financiación o en las estrategias de mercado de una empresa.

Cuando el número de posibles alternativas es finito, podemos resolver un problema de optimización mediante la comparación de alternativas, eliminando aquellas que son inferiores. El problema más difícil aparece cuando el número de alternativas es infinito. Aquí veremos una serie de técnicas, basadas en la diferenciación de funciones, que sirven para dificultades de este tipo.

1.1. Definiciones básicas. Existencia de soluciones

En un problema de optimización distinguimos tres elementos:

1) **El conjunto factible** o **conjunto de oportunidades**, que designaremos aquí con la letra X.

2) La **función objetivo**, una función definida en X y que toma valores numéricos, $f: X \to \mathbb{R}$, que es un índice que asociamos a cada elemento de X.

3) Un criterio de decisión: buscaremos aquel elemento o aquellos 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:

$$\max_{x} f(x) \quad \text{y} \quad \min_{d\ x} f(x)$$
$$\text{s.a. } x \in X \qquad\qquad \text{s.a. } x \in X,$$

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

Ejemplo 1.6. Recordemos el ejemplo 1.4., en el que nos planteábamos encontrar el rectángulo del área máxima de todos aquellos que tienen perímetro 1. Definimos x como la longitud de la base y asignamos a y la altura de un rectángulo. Entonces, el conjunto factible de nuestro problema de optimización es

$$X = \{(x, y) \in R^2 : x \geq 0, y \geq 0, y\ 2x + 2y = 1\}$$

y la función objetivo (el área) es $f(x, y) = xy$. En lugar de definir X de forma separada, solemos escribir este problema de optimización de la forma siguiente:

$$\begin{aligned} &\max_{(x,\, y)} xy \\ &\text{s.a. } 2x + 2y = 1 \\ &\qquad\quad x \geq 0 \\ &\qquad\quad y \geq 0 \end{aligned}$$

En esta formulación, el conjunto factible X está descrito por una serie de restricciones que imponemos a las variables de decisión.

Ejemplo 1.7. Recordemos ahora el ejemplo 1.5., en el que queremos encontrar aquel precio p que aumenta el ingreso $I(p) = 100p - 2p^2$. La variable de decisión es el precio p, la función objetiva es el ingreso $I(p)$, y el conjunto factible, todos los precios posibles. Por ejemplo, dado que María quiere cobrar por las caracolas y no pagar por ellas, podemos suponer que el conjunto factible son todos los números reales no negativos.

Ejercicio

1.1. Una empresa utiliza dos factores de producción y produce un único bien. Si usa cantidades x_1 y x_2 de los dos factores, obtiene una cantidad $f(x_1, x_2)$ del bien producido. Los precios de los factores de producción son 2 y 5 respectivamente, de modo que el coste de usar unas cantidades (x_1, x_2) de los factores es $2x_1 + 5x_2$. La empresa quiere encontrar la combinación (x_1, x_2) que reduce al mínimo el coste de producir 10 unidades del producto. Identificad las variables de decisión, la función objetivo y el conjunto factible.

Lo primero que debemos tener en cuenta es que un problema de optimización, por sencillo que sea, no necesariamente tiene solución.

Ejemplo 1.8. Consideremos el problema

$$\max_{x} f(x)$$
$$\text{s.a. } x \in \mathbb{N}$$

en el que N son los números naturales. En otras palabras, queremos encontrar el mayor número natural (1, 2, 3, 4...), y este problema no tiene solución.

Decimos que el problema

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

tiene solución si hay un cierto elemento x^* en el conjunto X de modo que se cumpla 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.

Decimos que el problema

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

tiene solución si hay un cierto elemento x^* en el conjunto X, de modo que se cumple que $f(x^*)$ £ $f(x)$, 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.

Un hecho importante que debemos observar es que, cuando un problema de maximización tiene solución, no hay necesariamente un único maximizador. Lo que sí que es necesariamente único es el valor máximo: si hubiese dos máximos diferentes, siempre podríamos quedarnos con el mayor de los dos; esto quiere decir que el otro no puede ser en realidad ningún máximo. El ejemplo siguiente lo ilustra.

Ejemplo 1.9. Sea $X = \{-2, -1, 0, 1\}$, y sea $f(x) = x^2$, para todo $x \in X$. Supongamos que queremos maximizar f sobre X. Claramente, el problema tiene solución: el maximizador es $x^* = -2$ y el valor máximo es $f(-2) = 4$.

Supongamos ahora que $X = \{-2, -1, 0, 1, 2\}$, y que $f(x) = x^2$, para todo $x \in X$. Ahora el problema tiene dos maximizadores, $x^* = -2$ y $x^{**} = 2$, y el valor máximo continúa siendo 4. Observamos que $f(2) = f(-2) = 4$.

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

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

Ejemplo 1.10. Unos cuantos recorridos

- Si $X = \{-2, -1, 0, 1, 2\}$ y $f(x) = x^2$, entonces $f(X) = \{0, 1, 2\}$.
- Si $X = [0, 2]$ y $f(x) = x^2$, entonces $f(X) = [0, 4]$.
- Si $X = [-2, 2]$ y $f(x) = x^2$, entonces $f(X) = [0, 4]$.
- Si $X = \mathbb{R}$ y $f(x) = e^x$, entonces $f(X) = (0, \infty)$.
- Si $X = (0,\infty)$ y $f(x) = \log(x)$, entonces $f(X) = \mathbb{R}$.

El siguiente lema resulta simplemente de interpretar lo que significa la definición que hemos dado, que se refiere a cuando un problema de maximización (o de minimización) tiene solución.

Lema

El problema

$$\begin{array}{ll} \underset{x}{max} & f(x) \\ \text{s.a.} & x \in X \end{array}$$

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

Lema

El problema

$$\begin{array}{ll} \underset{x}{min} & f(x) \\ \text{s.a.} & x \in X \end{array}$$

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 $a^* \geq a$. De forma análoga, 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, aunque no suficiente, para que un conjunto $A \subset \mathbb{R}$ tenga un máximo, es que A esté acotado de forma superior. Para ver que la condición no es suficiente, considerad el ejemplo siguiente: el intervalo semiabierto [0,1) es acotado superiormente, pero no tiene un máximo, ya que aunque el número 1 no está dentro del intervalo, cualquier número positivo menor que 1 sí que lo está (por ejemplo, 0,9999999, con tantos nueves como queráis).

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. Para ver que la condición no es necesaria, considerad el ejemplo siguiente: 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.

Unos cuantos ejemplos mostrarán diferentes tipos de inconvenientes que se pueden presentar en problemas de optimización.

Ejemplo 1.11. El recorrido no es acotado

Consideremos el problema

$$\begin{aligned} &\max_{x} \ x^2 \\ &\text{s.a.} \ \ x \in \mathrm{R}, \end{aligned}$$

Dado cualquier número M arbitrariamente grande, siempre podemos encontrar un $x \in \mathrm{R}$ tal que $x^2 > M$. Esto implica que el problema no tiene solución. Debéis notar que aquí el inconveniente es que el conjunto $f(X) = [0,\infty)$ no es acotado superiormente y, en consecuencia, no tiene un máximo. Por otro lado, observad que el recorrido es acotado en este caso por debajo, pero esto carece de importancia si queremos maximizar la función.

El hecho de restringir el conjunto factible a un intervalo acotado no soluciona necesariamente el problema.

Ejemplo 1.12. El recorrido es acotado, pero no tiene máximo

Consideremos

$$\begin{aligned} &\max_{x} \ x^2 \\ &\text{s.a.} \ \ x \in (0, 1), \end{aligned}$$

donde el recorrido es $f(X) = (0,1)$, que no tiene máximo. Podemos encontrar puntos $x \in X$ tales que $f(x)$ es arbitrariamente cercano a 1, pero no hay ningún punto que tenga 1 como imagen.

Ejemplo 1.13. El conjunto factible es acotado, pero el recorrido no lo es

Consideremos el problema

$$\begin{array}{ll} \max\limits_{x} & x^2 \\ \text{s.a.} & x \in [0, 1], \end{array}$$

en el que la función *f* viene definida por

$$f(x) = \begin{cases} \dfrac{1}{x}, & \text{si} \quad x > 0; \\ 0, & \text{si} \quad x = 0. \end{cases}$$

En este caso, dado cualquier número M arbitrariamente grande, siempre podemos encontrar un x suficientemente pequeño que cumpla $\frac{1}{x} > M$. Por lo tanto, el recorrido no es (superiormente) acotado y el problema no tiene solución.

El inconveniente en este ejemplo...

... reside en la falta de continuidad de la función *f*: si ésta hubiese sido continua, entonces el recorrido habría sido acotado.

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 los que una solución existe aunque aquellas condiciones no estén satisfechas. Los ejemplos que acabamos de discutir son ilustrativos del tipo de problemas que pueden presentarse cuando las condiciones (suficientes) del teorema siguiente son violadas.

Teorema de Weiertrass

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.

Debéis notar que este teorema expresa condiciones en términos del conjunto factible X y la función objetivo *f*, que son en general mucho más fáciles de verificar que requerimientos sobre el recorrido $f(X)$ (que es, según sabemos, lo que finalmente importa). Éste es un motivo adicional que hace que muchos problemas tengan solución aunque no cumplan las condiciones del teorema de Weiertrass. Otro motivo es que el teorema resulta tan válido para problemas de maximización como para problemas de minimización.

El motivo por el que este teorema es cierto es muy sencillo: siempre que los dos requerimientos se satisfagan, el recorrido $f(X)$ también es un conjunto cerrado y acotado (compacto) y, por lo tanto, tiene un máximo y un mínimo.

La utilidad del teorema es relativa, ya que nos indica que existe una solución del problema de optimización, pero no nos dice cómo la podemos encontrar.

Una aplicación del teorema de Weierstrass sería que, siempre que X sea un conjunto finito, cualquier problema de optimización que tenga X como conjunto factible tiene solución. Esto se debe al hecho de que todo conjunto finito es compacto, y que toda función definida sobre un conjunto finito es (trivialmente) continua sobre este conjunto de puntos (aunque no tiene por qué ser continua sobre un dominio mayor).

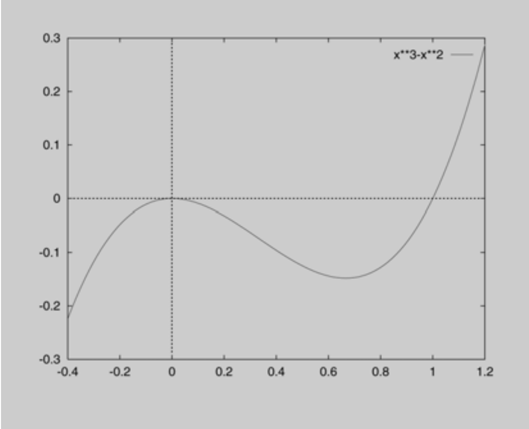
Un motivo importante por el que nos interesará tener un resultado general que nos diga cuándo existe una solución es que, si un problema no tiene solución pero nosotros aplicamos técnicas de resolución como si la tuviese, podemos obtener una respuesta que nos parezca coherente y no darnos cuenta del error que hemos cometido.

Ejemplo 1.14. Supongamos que queremos resolver el programa

$$\begin{array}{ll} \underset{x}{max} & x^3 - x^2 \\ \text{s.a.} & x \in \mathrm{R} \end{array}$$

y que sabemos que, cuando el problema tiene solución, la derivada de la función objetivo en el maximizador es igual a cero. Igualando la derivada de la función objetivo a cero, obtenemos la ecuación $3x^2 - 2x = 0$, que tiene por soluciones los puntos $x^* = 0$ y $x^{**} = \frac{2}{3}$. Aplicando el criterio de la derivada segunda, veremos que x^* es un maximizador local y que x^{**} es un minimizador local. Por lo tanto, el único candidato a maximizador es el punto $x^* = 0$. Si ahora concluimos que x^* es la solución del problema que habíamos planteado, estaremos cometiendo un error considerable, ya que en este caso, el recorrido no es acotado ni superior ni inferiormente.

Un vistazo a la gráfica de la función nos puede ayudar a ver por qué: la función no es acotada superior ni inferiormente.



Gráfica de la función $f(x) = x^3 - x^2$.

La función $f(x) = x^3 - x^2$...

... tiene un máximo y un mínimo locales, pero no tiene ningún máximo ni ningún mínimo globales.

Acabamos este apartado con un ejemplo que nos muestra que ninguna de las dos condiciones que se imponen en el teorema de Weierstrass es necesaria.

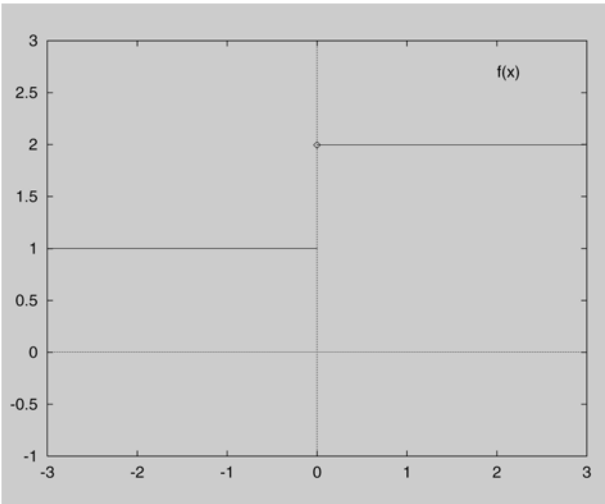
Ejemplo 1.15. Considerad el problema

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

en el que la función f viene definida por

$$f(x) = \begin{cases} 1, & \text{si} \quad x < 0; \\ 2, & \text{si} \quad x \geq 0. \end{cases}$$

Aquí el valor máximo es 2 y cualquier x en el intervalo $[0,\infty)$ es un maximizador, pero f no es continua y X no es acotado.



Una función constante a trozos

He aquí...

... una función con muchos maximizadores y muchos minimizadores, pero que no cumple las condiciones del teorema de Weiertrass.

1.2. Resolución de problemas. Algoritmos

En problemas de optimización no nos interesa sólo saber que hay una solución, sino encontrarla y describir sus propiedades. En este curso analizaremos fundamentalmente una serie de técnicas que se aplican cuando en los problemas de optimización se dan unas ciertas condiciones de diferenciabilidad que permiten describir la solución mediante un sistema de ecuaciones (o de desigualdades).

En problemas prácticos de optimización es difícil que estos requerimientos se satisfagan, por lo que probablemente la resolución del problema se hace con un ordenador. En este caso, tendremos que programar en el ordenador una secuencia de pasos que esperaremos que nos acaben conduciendo a la solución del problema o que nos den una buena aproximación al mismo.

Un algoritmo es un conjunto de reglas que describen los pasos que hay que seguir para buscar (o aproximar) la solución de un problema determinado.

Ejercicio

1.2. Sea $X = \{1,2... n\}$ un conjunto finito, y sea $f: X \to \mathbb{R}$ una función cualquiera. Describid (de modo que pudiéramos programarlo en un ordenador) un algoritmo que permita encontrar el máximo de la función f sobre X.

Con la (poca) información que se da en el problema anterior, todos los algoritmos que pueden definirse pasan por lo que podemos llamar enumeración de los elementos de X: básicamente, tenemos que ir tomando elementos de X y compararlos, y hasta que no los hayamos tomado todos no podremos saber con seguridad cuál es el máximo. Dicho en otras palabras, cuando X es finito, un problema de optimización siempre tiene solución, pero encontrarla puede ser muy costoso: n puede ser muy grande. A modo de ejemplo, supongamos que debemos hacer 2^{64} comparaciones: si nuestro ordenador puede hacer una comparación cada microsegundo, tardaría unos 5.000 siglos en obtener el resultado final.

El número 2^{64} ...

... está relacionado con aquella leyenda de un noble chino que preguntó a un sabio sobre lo que quería como pago de un servicio que éste le había hecho, y el sabio le pidió que tomase un tablero de ajedrez y pusiera un grano de trigo en la primera casilla, y que fuese doblando el número de granos en cada casilla sucesiva; el noble se fregó las manos, confiado, hasta que tuvo que llevar a cabo la promesa. El número total de granos que tendría que haber puesto es 2^{64}–1. ¿Podríais decir por qué?

En general, por lo tanto, nos interesará caracterizar y explotar propiedades que permitan simplificar los algoritmos de solución de problemas. De entre todas estas propiedades veremos en este curso la de diferenciabilidad y las de concavidad/convexidad.

El gran éxito de la teoría de la optimización ha sido el hecho de que se descubrió un algoritmo eficiente para solucionar problemas de programación lineal, y que muchas situaciones prácticas que se presentan en diferentes campos se pueden resolver con programas lineales. Por ejemplo, el modo de determinar los horarios de una compañía aérea. ❗

El gran reto en la rama algorítmica de la teoría de la optimización hoy en día es el de desarrollar algoritmos eficientes para la resolución de problemas no lineales. En este curso no veremos esta vertiente algorítmica de la teoría de la optimización, algo de lo que se encarga en mayor medida otro curso llamado *Investigación operativa*.

Una anécdota...

... que ilustra lo que en su momento significó el desarrollo de la programación lineal es el hecho de que, durante la Segunda Guerra Mundial, las técnicas de resolución de programas lineales en Estados Unidos eran un secreto militar.

1.3. Diferenciación y optimización

A partir de ahora supondremos que nuestra variable de decisión es un vector n–dimensional, es decir, que X es un subconjunto de $\mathbb{R}^n$. También supondremos en general que la función objetivo f es diferenciable una vez o más.

En este caso, la herramienta fundamental para encontrar la solución del problema es la diferenciación. Como sabéis, la **diferenciación** es una forma de aproximar localmente una función cualquiera a un cierto punto mediante una función lineal; esto nos permite extraer numerosas conclusiones sobre el comportamiento (a nivel local) de la función original; en particular, si la función tiende a crecer o a decrecer en alguna dirección alrededor del punto de partida de la aproximación.

La característica que hace que la diferenciación sea tan importante para resolver problemas de optimización es la relación que hay entre la derivada de una función (univariante) y el hecho de que ésta sea creciente o decreciente.

Dedicaremos este apartado a recordar brevemente la relación entre la derivada y el crecimiento de una función.

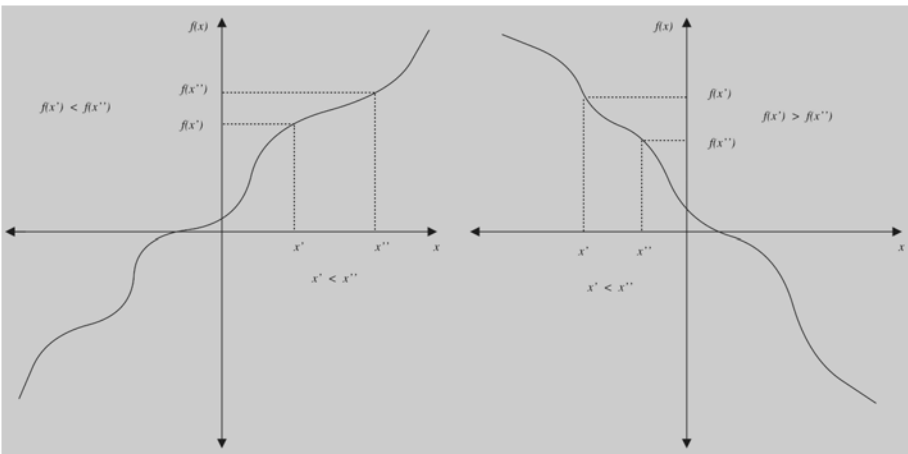
Sea I un intervalo de números reales (donde no descartamos que $I = \mathbb{R}$), y sea $f : I \rightarrow \mathbb{R}$ una función.

Decimos que la función f es **creciente** si $x' \leq x''$ implica que $f(x') \leq f(x'')$.

De forma análoga, f es decreciente si $x' \leq x''$ implica que $f(x') \geq f(x'')$.

Decimos que la función f es **estrictamente creciente** si $x' < x''$ implica que $f(x') < f(x'')$. De forma análoga, f es **estrictamente decreciente** si $x' < x''$ implica que $f(x') > f(x'')$.

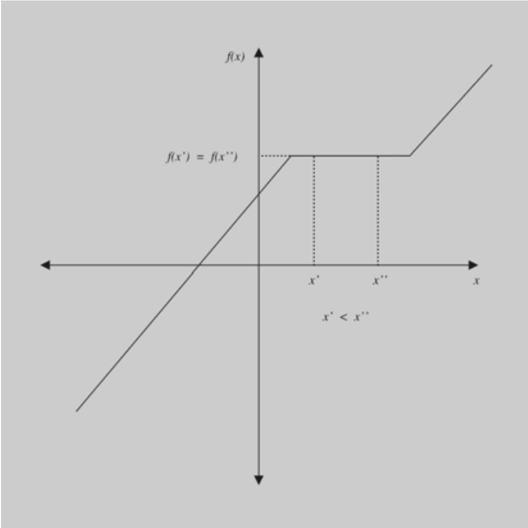
En otras palabras, una función es estrictamente creciente cuando mantiene la disposición de los puntos de dominio, y estrictamente decreciente cuando la invierte.



Una función estrictamente creciente

Una función estrictamente decreciente

La diferencia entre una función creciente en general y una creciente estrictamente es que la primera puede tener algunos sectores planos.



Crecimiento pero no estricto

Observad que...

...$x' < x''$ pero $f(x') = f(x'')$. Esto no pasaría si la función fuese estrictamente creciente.

Ejemplo 1.16. Toda función constante es tanto creciente como decreciente, pero no lo es estrictamente en ningún sentido.

Recordemos el siguiente teorema para funciones diferenciables de una variable.

Teorema del valor medio

Sea $f: [a, b] \rightarrow \mathbb{R}$ una función continua y diferenciable en (a, b). Entonces existe un número $c \in (a, b)$ tal que

$$f(b) - f(a) = f'(c)\,(b - a).$$

Los resultados que aparecen a continuación son consecuencia del teorema del valor medio.

Proposición 1. Supongamos que $f: I \rightarrow \mathbb{R}$ es diferenciable. Entonces tenemos que:

- f es **creciente** en I si, y sólo si, $f'(x) \geq 0$ para todo $x \in I$.
- f es **decreciente** en I si, y sólo si, $f'(x) \leq 0$ para todo $x \in I$.

Ejemplo 1.17. Si f es una función constante, entonces su derivada es siempre 0. Por lo tanto, f es creciente. Los mismos motivos nos llevan a afirmar que f es decreciente. De hecho, las funciones constantes son las únicas que pueden ser crecientes y decrecientes al mismo tiempo (no en sentido estricto, claro).

La proposición anterior nos da una caracterización completa de funciones crecientes o decrecientes, sobre la base de su derivada. A continuación veremos otro resultado que, aunque no siempre es decisorio, es extremamente útil.

Proposición 2. Supongamos que $f : I \to \mathrm{R}$ es diferenciable. Entonces tenemos que:

- Si $f'(x) > 0$ para todo $x \in I$, entonces f es estrictamente creciente en I.
- Si $f'(x) < 0$ para todo $x \in I$, entonces f es estrictamente decreciente en I.

Se da un caso bastante excepcional que es importante destacar: una función puede ser estrictamente creciente o decreciente aunque su derivada sea 0 en algún punto. Sin embargo, observamos que esto sólo puede suceder en puntos aislados: si la derivada fuese cero en un intervalo entero, entonces la función sería constante en este intervalo y, por lo tanto, no podría ser creciente ni decreciente en sentido estricto. Un ejemplo de ello es la función $f(x) = x^3$, que es estrictamente creciente; su derivada $f'(x) = 3x^2$ es estrictamente positiva en todos los lugares, excepto en el punto $x = 0$, donde vale 0. En general, sin embargo, es conveniente asociar intuitivamente una función estrictamente creciente con una derivada positiva, y otra estrictamente decreciente con una derivada negativa.

Ejercicio

1.3. En cada uno de los casos que aparece a continuación tenemos que $f : \mathrm{I} \to \mathrm{R}$. Decid si la función es creciente o decreciente sobre el intervalo I en el que está definida; si vuestra respuesta es afirmativa, decid si lo es en sentido estricto o no. Indicad también en cada caso cuál es el recorrido J de la función:

$J = f(I) = \{ y : y = f(x) \text{ para algún } x \in I \}$.

Podéis recurrir a la ayuda del Gnuplot cuando tengáis alguna duda.

a) $f(x) = x$, $I = \mathrm{R}$.
b) $f(x) = -x$, $I = \mathrm{R}$.
c) $f(x) = 2$, $I = \mathrm{R}$.
d) $f(x) = 5 + 7x$, $I = \mathrm{R}$.
e) $f(x) = 5 - 7x$, $I = \mathrm{R}$.
f) $f(x) = x^2$, $I = \mathrm{R}$.
g) $f(x) = x^2$, $I = [0,\infty)$.
h) $f(x) = -x^3$, $I = \mathrm{R}$.
i) $f(x) = e^x$, $I = \mathrm{R}$.
j) $f(x) = e^{-2x}$, $I = \mathrm{R}$.
k) $f(x) = \log x$, $I = (0,\infty)$.
l) $f(x) = \log(1 + x^2)$, $I = \mathrm{R}$.
m) $f(x) = \log(1 + x^2)$, $I = [0, \infty)$.
n) $f(x) = \frac{1}{x}$, $I = (0, \infty)$
o) $f(x) = \sqrt{x}$, $I = [0, \infty)$
p) $f(x) = \frac{1}{1 + x^2}$, $I = \mathrm{R}$
q) $f(x) = \operatorname{sen} x$, $I = \mathrm{R}$

r) $f(x) = sen x, \quad I = \left(0, \frac{\pi}{2}\right).$

s) $f(x) = cos x, I = \mathbb{R}.$

t) $f(x) = cos x, \quad I = \left(0, \frac{\pi}{2}\right).$

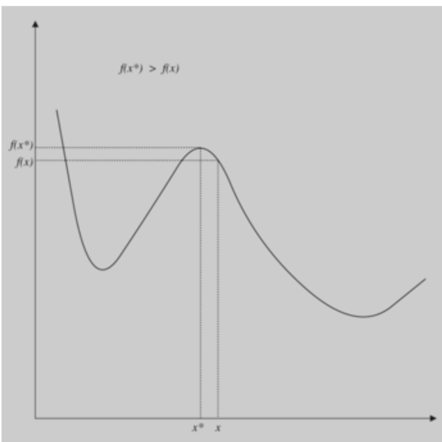
Es conveniente que os fijéis en que, dado que la diferenciación es una herramienta de aproximación local de una función, todos los resultados que obtengamos a partir de esta herramienta tendrán sólo un carácter local y, por lo tanto, debemos tener mucho cuidado en su interpretación. En particular, en el estudio de problemas de optimización nos interesa introducir los conceptos siguientes:

Sea $X \subset \mathbb{R}^n$ y $f: X \to \mathbb{R}$.

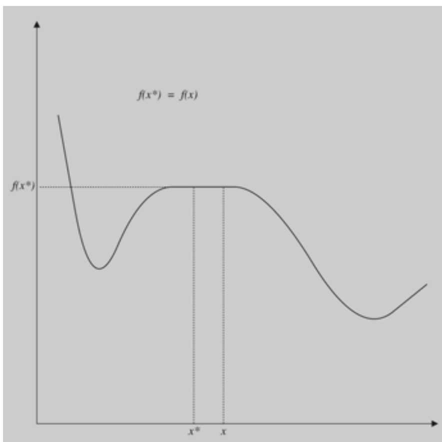
Decimos que f alcanza un **máximo local** en un punto $x^* \in X$ si hay un cierto entorno de x^* tal que $f(x^*) \geq f(x)$, para todo punto x de aquel entorno. Si la desigualdad es estricta para todo punto del entorno diferente de x^*, entonces decimos que el máximo local es **estricto**.

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

En un **máximo local estricto**, el valor de la función en el maximizador es estrictamente mayor que en todos los puntos que lo rodean. En cambio, en un **máximo local no estricto**, la función toma un valor mayor o igual que en los puntos de alrededor.



Máximo local estricto



Máximo local no estricto

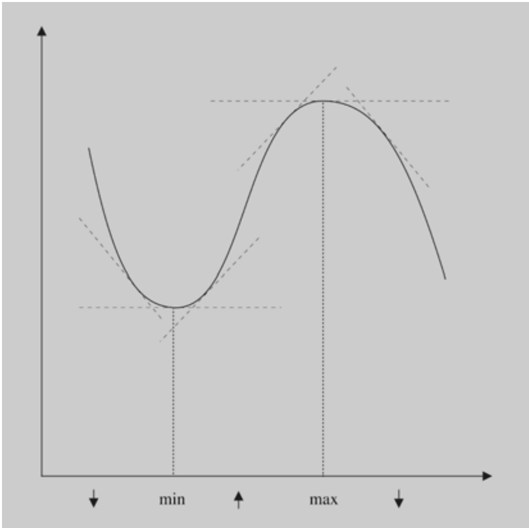
Para hacer la distinción entre los maximizadores o los minimizadores locales y los maximizadores o los minimizadores que solucionan un problema de optimización, en ocasiones añadiremos a estos últimos la etiqueta de **globales**.

Algunos autores hablan de extremos **relativos**, para designar los que nosotros llamamos locales. Nosotros preferimos la terminología de *locales* frente a *globales*, más que la de *relativos* frente a *absolutos*, porque creemos que la primera es más descriptiva (además, su uso se está imponiendo cada vez más).

Utilizaremos el término *extremo*...

... para referirnos a un punto que puede ser tanto un máximo como un mínimo.

Ya habéis visto con anterioridad la herramienta básica de detección de máximos y mínimos locales de funciones diferenciables; consiste en que en un extremo local la derivada de la función debe ser igual a cero. Acabamos de ver que una función crece cuando su derivada es positiva y decrece cuando ésta es negativa. Dado que en un mínimo local la función pasa de decreciente a creciente, la derivada no puede ser ni negativa ni positiva, así que debe ser cero. El mismo argumento se aplica a los máximos locales, en los que la función pasa de creciente a decreciente.

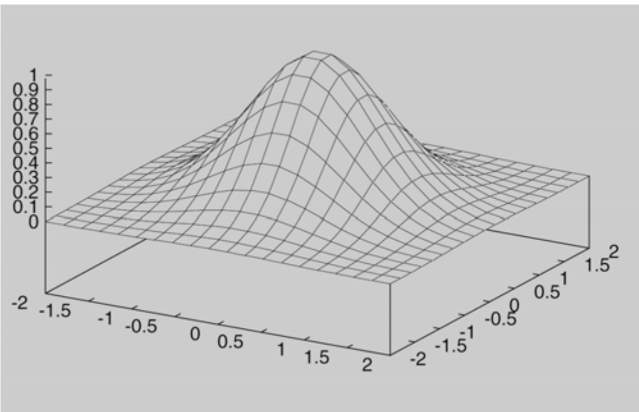


Máximos y mínimos

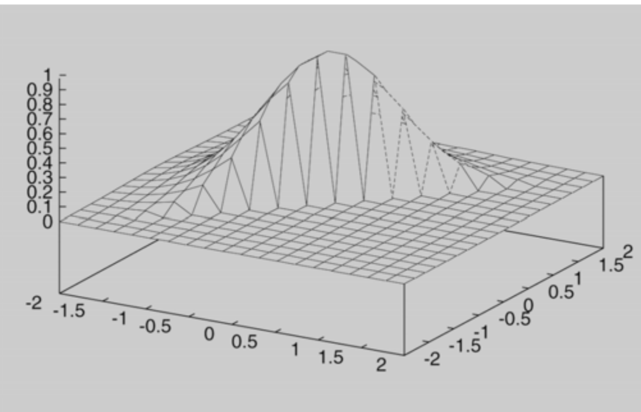
En un mínimo local,...

... la función pasa de decreciente a creciente. En un máximo local, la función pasa de creciente a decreciente. Por lo tanto, en ambos puntos la derivada debe ser cero.

En el caso de las funciones multivariantes, un resultado paralelo es cierto; en un extremo local todas las derivadas parciales deben ser cero. El motivo es que, si la función tiene un máximo local en un punto, entonces cualquier sección vertical también debe tener un máximo local en el mismo punto, y las derivadas parciales no son más que las derivadas de las secciones verticales en las direcciones de los ejes.



Máximo en dos variables



Sección vertical de un máximo

Ejercicio

1.4. Haced la gráfica con Gnuplot de la función de dos variables

$$f(x,y)= e^{-x^2-y^2}$$

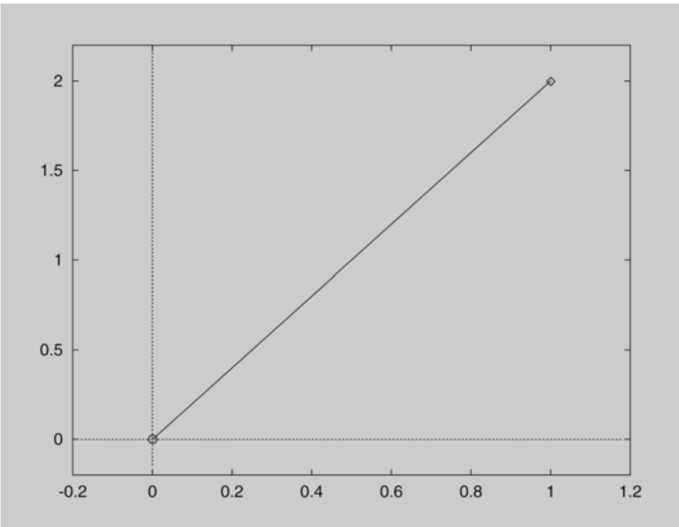
y comprobad que se corresponde con la que mostramos arriba. En particular, la función tiene un valor máximo en el punto (0, 0). A continuación realizad secciones verticales que pasen por el origen y comprobad que todas tienen un máximo en el mismo punto.

1.4. Tipos de programas de optimización

El criterio de que en un extremo las derivadas parciales deben ser cero se basa, sin embargo, en un supuesto muy importante: este extremo debe ser un punto interior; es decir, en cualquier dirección a partir del punto debe haber otros que también sean factibles. Cuando no sea éste el caso, diremos que estamos tratando un problema con restricciones, por lo que tendremos que desarrollar herramientas específicas para detectar los extremos locales.

Ejemplo 1.18. Un problema con restricciones

Supongamos que estamos buscando los extremos locales de la función $f(x) = 2x$, sobre el conjunto $X = [0, 1]$. Tenemos que $f'(x) = 2$, para todo x. Esto parecería indicar que la función no tiene máximos ni mínimos. Sin embargo, un vistazo a la gráfica tendría que ser suficiente para apreciar que f tiene un mínimo (global) en el punto $x = 0$, y un máximo (global) en el punto $x = 1$.



Máximo y mínimo restringidos

Dado que la función...

... es estrictamente creciente sobre [0,1], tiene un mínimo en el extremo inferior del intervalo y un máximo en el extremo superior.

Observemos que, si hubiese otros puntos a la izquierda de $x = 0$, entonces éste no sería un minimizador; pero dado que estamos imponiendo la restricción $x \geq 0$, entonces el hecho de que la derivada sea positiva no impide que este punto sea un minimizador.

En particular, si el conjunto factible es un **conjunto abierto**, todos sus puntos son interiores y, por lo tanto, cualquier maximizador o minimizador debe satisfacer la condición de que las derivadas parciales de la función objetivo se anulen. Este efecto motiva la siguiente definición.

> Un problema de **optimización sin restricciones** es aquél en el que el conjunto factible es un conjunto abierto.

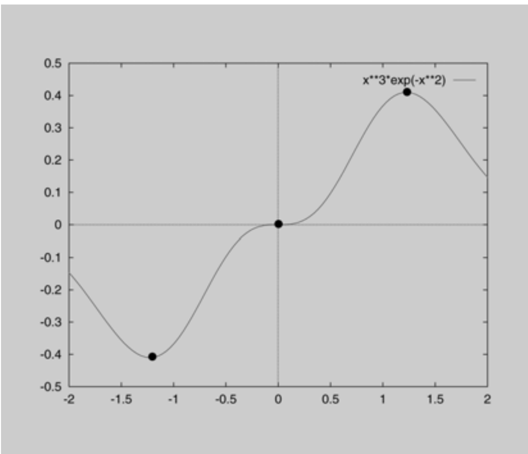
Ya hemos visto anteriormente que, además de los maximizadores y los minimizadores locales del problema, puede haber otros puntos en los que las derivadas parciales son iguales a cero.

Ejemplo 1.19. Extremos y puntos de inflexión

Sea $f(x) = f'(x) = x^3e^{-x^2}$. Al igualar la derivada a cero obtenemos:

$$f'(x) = e^{-x^2}(3x^2 - 2x^4) = 0 \to 3x^2 = 2x^4,$$

donde hay tres soluciones: $x^* = \sqrt{\frac{3}{2}}, x^{**} = -\sqrt{\frac{3}{2}}$ y $x^{***} = 0$. Podéis comprobar ahora que la función es monótona en los extremos: si $x > x^*$ o $x < x^{**}$, entonces $3x^2 < 2x^4$ y, por lo tanto, f es estrictamente decreciente. Además, el único punto en el que f corta el eje de las x tiene lugar cuando $x = 0$. Teniendo en cuenta lo que acabamos de observar, la porción relevante de la gráfica de la función nos ayudará a ver que la primera solución es el maximizador (global), la segunda el minimizador (global), y la tercera no es ni maximizador ni minimizador, sino un punto de inflexión.



Gráfica de $x^3e^{-x^2}$

En la gráfica...

... podemos visualizar los tres puntos críticos, que son, de izquierda a derecha, un minimizador, un punto de inflexión y un maximizador.

Las técnicas que usaremos en todos los problemas de optimización que estudiaremos (ya sean sin restricciones o con restricciones) tienen un rasgo común: basándonos en la diferenciación, formularemos una serie de relaciones que necesariamente debe cumplir cualquier punto que solucione el problema de optimización (si es que hay alguno que lo haga). Denominaremos estas relaciones **condiciones de primer orden**, ya que están basadas en las derivadas de primer orden. Los puntos que solucionan las condiciones de primer orden son los **puntos críticos**. Entre estos puntos críticos puede haber algunos que solucionen nuestro problema de optimización, y muchos otros que no lo consigan.

En los puntos críticos que no son extremos, la función debe crecer necesariamente a lo largo de algunas direcciones factibles y debe decrecer a lo largo de otras. Cuando no hay restricciones, y por lo tanto todas las direcciones son factibles, un punto crítico que no es ningún extremo es un **punto de silla** (lo cual no quiere decir que necesariamente se parezca a una silla de montar).

Ejercicios

1.5. Un punto de silla literal

Demostrad que (2, 3) es un punto crítico de la función $f(x–y) = –(x–2)^2 + (y–3)^2$. Usad Gnuplot para generar la gráfica de la función alrededor de este punto, y comprobad a partir de la gráfica que obtengáis que se trata de un punto de silla.

1.6. Un punto de silla sin silla

También puede haber puntos de silla con los cuales nos sería difícil montar a caballo. Aquí tenemos un ejemplo de ello. Demostrad, a partir de su derivada, que (2, 3) es un punto crítico de la función $f(x,y) = (x–2)^3 + (y–3)^3$. Usad Gnuplot para generar la gráfica de la función alrededor de este punto. Comprobad en la gráfica que, a partir del punto, la función crece en unas direcciones y decrece en otras y que, por lo tanto, se trata de un punto de silla, aunque la forma de la gráfica no se parezca demasiado a una silla de montar.

Una cuestión que se nos presenta es la de saber cuándo un punto crítico es un maximizador local, un minimizador local o un punto de silla. En el caso de una sola variable, el criterio que se usa es el de observar si los puntos críticos satisfacen las **condiciones de segundo orden**, que son restricciones sobre la segunda derivada: este criterio será determinante siempre que la segunda derivada tenga un signo estricto, y en caso contrario debemos continuar investigando. En el caso de más de una variable, tenemos criterios similares, en los que en lugar de la segunda derivada debemos considerar la matriz de derivadas parciales segundas. Más adelante ya veremos todos estos conceptos con detalle.

Como hemos visto, un problema de optimización cambia radicalmente cuando hay puntos desde los cuales ciertas direcciones son factibles y otras no lo son. En un caso así, ya no es cierto que todas las derivadas parciales de un maximizador o minimizador local deban ser necesariamente iguales a cero. Clasificaremos los problemas con restricciones en dos categorías, según si las restricciones están formadas por igualdades o por desigualdades.

Un problema de **optimización con restricciones de igualdad** es aquél en el que el conjunto factible está formado por todos aquellos puntos que cumplen una serie de restricciones en forma de igualdad (es decir, que solucionan un sistema de ecuaciones).

Un problema de **optimización con restricciones de desigualdad** es aquél en el que el conjunto factible está formado por todos aquellos puntos que cumplen una serie de restricciones en forma de desigualdad (y tal vez también de igualdad).

Los problemas...

... con restricciones de igualdad resultarán, en esencia, muy parecidos a los problemas sin restricciones.

Ejemplo 1.20. Restricciones de igualdad

El problema

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

puede ser transformado en el problema (sin restricciones) con una sola variable

$$\begin{aligned} \min_{x} \ & x^2 + (2-x)^2 \\ \text{s.a.} \ & x \in \mathbb{R}, \end{aligned}$$

pero cuando tratemos este tipo de problemas veremos que hay una técnica más poderosa que nos permite ahorrarnos esta sustitución y encontrar directamente condiciones que cualquier minimizador local debe satisfacer.

Cuando tengamos problemas con restricciones de desigualdad, el asunto será un poco más complicado, ya que debemos ser capaces de determinar qué direcciones son factibles y cuáles no lo son.

Ejemplo 1.21. Restricciones de desigualdad

El problema

$$\begin{aligned} \max_{x} \ & x^2 \\ \text{s.a.} \ & x \geq 0 \\ & x \leq 1 \end{aligned}$$

tiene un maximizador global que es $x^* = 1$. Desde $x^* = 1$ es factible moverse hacia el 0, pero no hacia el otro lado. En consecuencia, la condición que debe satisfacer la derivada de la función objetivo en 1 para que f alcance un valor máximo no será la igualdad a cero, sino una desigualdad: siempre que la derivada sea positiva, esto querrá decir que, yendo hacia 0, el valor de la función objetivo disminuirá, por lo que nos encontraremos sobre un maximizador local.

En un problema con restricciones de desigualdad diremos que una restricción es **efectiva** en un punto si este punto satisface la restricción en forma de igualdad. Por ejemplo, la restricción $x \leq 1$ es efectiva en el punto $x^* = 1$, pero la restricción $x \geq 0$ no lo es. La técnica que nos permitirá encontrar la solución de estos problemas consiste en escribir una serie de relaciones que permiten investigar qué restricciones son efectivas en el óptimo y cuáles no lo son.

En todos los tipos de problemas de optimización que acabamos de ver (con restricciones o sin éstas), la técnica usada para tratar de resolver el problema es la diferenciación, porque nos permite transformar el problema de encontrar un punto óptimo en el problema de resolver un sistema de ecuaciones. Sin embargo, siendo la diferenciación una aproximación local, los resultados que obtenemos con su aplicación también son de carácter local, mientras que en los problemas de optimización de los que partimos lo que se busca es una solución global.

Las herramientas que nos permitirán inferir el carácter global de las soluciones locales que hemos encontrado son dos:

a) El requerimiento de que el conjunto factible sea convexo.

b) El requerimiento adicional de que la función objetivo tenga la curvatura (concavidad o convexidad) adecuada según si queremos maximizar o minimizar.

En la parte de cálculo univariante habéis visto que una función (diferenciable dos veces) es convexa si, y sólo si, $f''(x) \geq 0$ en cada x, y que es cóncava si la desigualdad opuesta se cumple. Este resultado lo generalizaremos en el caso de varias variables utilizando matrices en lugar de números.

Ejercicios

1.7. Considerad el problema

$$\begin{array}{ll} \underset{x}{max} & 1 - x^2 \\ \text{s.a.} & x \in \mathrm{R}, \end{array}$$

a) Demostrad que el único punto crítico es $x^* = 0$.

b) Usando Gnuplot, elaborad una gráfica de la función alrededor de $x^* = 0$.

c) Teniendo en cuenta los signos de $f'(x)$, probad que f es creciente a la izquierda de 0 y decreciente a su derecha y que, por lo tanto, 0 es un maximizador global.

d) Una forma alternativa de observar que 0 es la solución del problema es probar directamente que, para todo $x \in \mathrm{R}$, se satisface $f(0) = 1 \geq 1 - x^2 = f(x)$. Hacedlo y completad el argumento.

1.8. Considerad el problema

$$\begin{array}{ll} \underset{x}{min} & x \ \log x \\ \text{s.a.} & x \in (0, 1) \end{array}$$

en el que $x \in (0,1)$ quiere decir que $0 < x < 1$; por lo tanto, el conjunto factible es abierto.

a) Usando Gnuplot, elaborad una gráfica de la función.

b) Demostrad que el único punto crítico es $x^* = \dfrac{1}{e}$.

c) Probad que f es decreciente a la izquierda de x^* y creciente a su derecha y que, por lo tanto, x^* es un minimizador global.

1.9. Considerad el problema

$$\max_x \quad \frac{1}{1+x^2}$$
$$\text{s.a.} \quad x \in \mathrm{R}$$

a) Demostrad que el único punto crítico es $x^* = 1$

b) Probad que f es creciente a la izquierda de 1 y decreciente a su derecha y que, por lo tanto, 1 es un maximizador global.

c) Probad directamente que, para todo $x \in \mathrm{R}$,

$$\frac{1}{1+x^2} \leq 1 .$$

d) Usando Gnuplot, haced una gráfica de la función alrededor de $x^* = 1$.

1.5. Solucionario

1.1. Variables de decisión: cantidades de los factores de producción (x_1 y x^2).

Función objetivo: coste de producción ($2x_1 + 5x^2$).

Conjunto factible: todas aquellas combinaciones de factores, (x_1,x^2), que permiten producir 10 unidades:

$$\{(x_1,x_2):f(x_1,x_2)\geq 10\}$$

1.2. El algoritmo consiste en ir comparando los números sucesivamente. Tenemos un candidato a máximo que vamos cambiando a medida que encontramos números mayores. Empezamos tomando $f(1)$ como candidato máximo. Entonces lo comparamos con $f(2)$, y el mayor de los dos es el nuevo candidato a máximo, que comparamos con $f(3)$, y así sucesivamente. El resultado de comparar el último candidato con $f(n)$ nos dará el máximo de la función.

1.3.

a) $f(x) = x$, $I = \mathrm{R}$: estrictamente creciente, ya que, para todo x, se cumple $f'(x) = 1>0$.

b) $f(x) = -x$, $I = \mathrm{R}$: estrictamente decreciente, ya que, para todo x, se cumple $f'(x) = -1<0$.

c) $f(x) = 2$, $I = \mathrm{R}$: constante (es decir, es al mismo tiempo débilmente creciente y débilmente decreciente).

d) $f(x) = 5 + 7x$, $I = \mathrm{R}$: estrictamente creciente, ya que, para todo x, se cumple $f'(x) = 7>0$.

e) $f(x) = 5- 7x$, $I = \mathrm{R}$: estrictamente decreciente, ya que, para todo x, se cumple $f'(x) = -7<0$.

f) $f(x) = x^2$, $I = \mathrm{R}$: no es creciente ni decreciente, ya que, por ejemplo, tenemos $f'(-1) = -2<0<2 = f'(1)$.

g) $f(x) = x^2$, $I = [0,\infty)$: estrictamente creciente, ya que $f'(x) = 2x>0$, para todo $x>0$.

h) $f(x) = -x^3$, $I = \mathrm{R}$: estrictamente creciente, ya que $f'(x) = 3x^2>0$, para todo $x\neq 0$.

i) $f(x) = e^x$, $I = \mathrm{R}$: estrictamente creciente, ya que, para todo x, se cumple $f'(x) = e^x>0$.

j) $f(x) = e^{-2x}$, $I = \mathrm{R}$: estrictamente decreciente, ya que, para todo x, se cumple $f'(x) = -2e^{-2x}<0$.

k) $f(x) = \log x$, $I = (0,\infty)$: estrictamente creciente, ya que, para todo x, se cumple $f'(x) = 1/x>0$.

l) $f(x) = log(1 + x^2)$, $I = \mathbb{R}$: No es creciente ni decreciente, ya que, por ejemplo, tenemos $f'(-1) = -1 < 0 < 1 = f'(1)$.

m) $f(x) = log(1 + x^2)$, $I = [0,\infty)$: estrictamente creciente, ya que $f'(x) = \dfrac{2x}{1+x^2} > 0$, para todo $x>0$.

n) $f(x) = \dfrac{1}{x}$, $I = (0, \infty)$ estrictamente decreciente, ya que, para todo x, se cumple $f'(x) = -x^{-2} < 0$.

o) $f(x) = \sqrt{x}$, $I = [0, \infty)$:estrictamente creciente, porque $f'(x) = \dfrac{1}{2\sqrt{x}} > 0$, para todo $x > 0$

p) $f(x) = d\dfrac{1}{1+x^2}$, $I = \mathbb{R}$. No es creciente ni decreciente, ya que, por ejemplo, tenemos $f'(1) = -1/2 < 0 < 1/2 = f'(-1)$.

q) $f(x) = sen\ x$, $I = \mathbb{R}$: no es creciente ni decreciente, ya que, por ejemplo, tenemos $f'(\pi) = -1 < 0 < 1 = f'(0)$.

r) $f(x) = \operatorname{sen} x$, $I = \left(0, \dfrac{\pi}{2}\right)$: estrictamente creciente, porque $f'(x) = cos x > 0$, para todo $0 \leq x < \dfrac{\pi}{2}$.

s) $f(x) = cos x$, $I = \mathbb{R}$: no es creciente ni decreciente, ya que, por ejemplo, tenemos $f'(\pi/2) = -1 < 0 < 1 = f'(3\pi/2)$.

t) $f(x) = \cos x$, $I = \left(0, \dfrac{\pi}{2}\right)$: estrictamente decreciente, porque $f'(x) = -sin x < 0$, para todo $0 < x < \pi/2$.

1.4. La gráfica de la función la obtenemos haciendo:

```
gnuplot> f(x,y) = exp (–x**2–y**2)
gnuplot> set isosamples 20
gnuplot> set hidden3d
gnuplot> splot [–2:2] [–2:2] f(x,y)
```

Para obtener el gráfico del corte vertical, hacemos:

```
gnuplot> g(x,y) = (x<y) ? f(x,y) : 0
gnuplot> splot [–2:2] [–2:2] g(x,y)
```

1.5. Dada $f(x,y) = -(x-2)^2 + (y-3)^2$, sus derivadas parciales son

$$\frac{\partial f}{\partial x} = -2(x-2)$$
$$\frac{\partial f}{\partial y} = 2(y-3)$$

Y es obvio que estas derivadas parciales son cero en el punto (2, 3). Cuando fijamos $y = 3$, el corte vertical de la función f que obtenemos es $g(x) = -(x-2)^2$, que tiene un máximo en el punto $x = 2$. Por otro lado, si fijamos $x = 2$, el corte vertical de la función f que obtenemos es $h(y) = (y-3)^2$, que tiene un mínimo en el punto $y = 3$. Por lo tanto, se trata de un punto de silla.

Con Gnuplot haremos:

```
gnuplot> f(x,y) = – (x–2)**2 + (y–3)**2
gnuplot> set hidden3d
gnuplot> splot [1:3] [2:4] f(x,y)
```

Si lo queremos ver mejor, podemos ir cambiando la perspectiva usando la instrucción set view.

1.6. Dada $f(x,y) = (x - 2)^3 + (y - 3)^3$, sus derivadas parciales son

$$\frac{\partial f}{\partial x} = 3(x-2)^2$$
$$\frac{\partial f}{\partial y} = 3(y-3)^2$$

Y es obvio que estas derivadas parciales son cero en el punto (2, 3). Cuando fijamos $y = 3$, el corte vertical de la función f que obtenemos es $g(x) = (x - 2)^3$, que es una función estrictamente creciente en x, por lo que en el punto (2, 3) la función f no puede tener un máximo ni un mínimo. Por definición, (2, 3) es un punto de silla.

Con Gnuplot haremos:

```
gnuplot> f(x, y) = (x – 2)**3 + (y – 3)**3
gnuplot> set hidden3d
gnuplot> splot [1:3] [2:4] f(x,y)
```

Podemos observar que no hay ninguna semejanza con una silla de montar.

1.7 Queremos resolver

$$\max_x \; 1 - x^2$$
$$\text{s.a.} \quad x \in \mathrm{R}$$

Sea $f(x) = 1–x^2$, entonces los puntos críticos solucionan la ecuación

$$f'(x) = -2x = 0 \Rightarrow x = 0$$

Además, podemos ver que f pasa de ser creciente a ser decreciente en $x = 0$, por lo que se trata de un máximo (global):

$$x<0 \Rightarrow f'(x) > 0, \quad x > 0 \Rightarrow f'(x) < 0$$

Alternativamente, comprobamos que, para cada x, $f(0) = 1 \geq 1-x^2 = f(x)$, ya que $x^2 \geq 0$.

1.8. Tenemos

$$\min_x \; x \log x$$
$$\text{s.a.} \quad x \in (0, 1)$$

Sea $f(x) = x \log x$. En este caso, los puntos críticos deben cumplir:

$$f'(x) = 1 + \log(x) = 0 \Rightarrow x = e^{-1} \in (0,1)$$

También tenemos

$$x < 1/e \Rightarrow f'(x) < 0, \; x > 1/e \Rightarrow f'(x) > 0$$

expresión que significa que $1/e$ es un minimizador global.

1.9. Tenemos

$$\max_x \; \frac{1}{1+x^2}$$
$$\text{s.a.} \quad x \in \mathrm{R}$$

Sea $f(x) = 1/(1 + x^2)$. Para obtener los puntos críticos, hacemos:

$$f'(x) = \frac{2x}{(1+x^2)^2} = 0 \Rightarrow x = 0$$

Además, podemos ver que f pasa de ser creciente a decreciente en $x = 0$, por lo que se trata de un máximo (global):

$$x < 0 \Rightarrow f'(x) > 0, \quad x > 0 \Rightarrow f'(x) < 0$$

Alternativamente, tenemos que, para todo x, $f(0) = 1 \geq f(x)$, ya que $1 + x^2 \geq 1$ para todo x real.

2. Optimización sin restricciones

2.1. Condiciones de primer orden

Hemos definido un problema de optimización sin restricciones como aquél en el que la función objetivo f es diferenciable una vez o más y, adicionalmente, el conjunto X es abierto, es decir, cada punto factible está rodeado de otros puntos que también lo son.

Ejemplo 2.1. Cualquier problema en el que el conjunto factible está formado por todo el espacio n-dimensional (con $n \geq 1$) es un problema sin restricciones. Por ejemplo:

$$\begin{array}{ll} \max\limits_{(x,y)} & e^{-x^2} - y^4 \\ \text{s.a.} & (x,y) \in \mathbb{R}^2. \end{array}$$

Ejemplo 2.2. El problema

$$\begin{array}{ll} \max\limits_{x} & x \log(x) \\ \text{s.a.} & x \in (0,1). \end{array}$$

es un problema sin restricciones, de acuerdo con la definición que acabamos de dar. El motivo es que si un punto satisface $0<x<1$, entonces hay un cierto número $e>0$ tal que el intervalo $(x-e, x+e)$ está enteramente contenido dentro del conjunto factible. (Por ejemplo, podríamos tomar e como el más pequeño de los números $\frac{x}{2}$ y $\frac{1-x}{2}$). Esto quiere decir que el punto x está rodeado de otros puntos factibles en cualquier dirección.

Ejemplo 2.3. El problema

$$\begin{array}{ll} \min\limits_{(x,y)} & 3x^2 + 2y^2 \\ \text{s.a.} & x^2 + y^2 < 1 \end{array}$$

es otro problema sin restricciones, de acuerdo con la definición que hemos dado más arriba. El motivo es que, al ser la restricción de la forma $g(x,y) = x^2 + y^2 < 1$, en el que $g(x,y)$ es una función continua, cualquier punto que satisfaga la desigualdad está necesariamente rodeado en cualquier dirección de otros puntos que también la satisfacen (esto es cierto siempre que tenemos una función continua cuyos valores satisfacen una desigualdad estricta).

Ejemplo 2.4. El problema

$$\begin{array}{ll} \min\limits_{x} & e^x \\ \text{s.a.} & x > 0 \end{array}$$

es un problema sin restricciones, de acuerdo con nuestra definición, porque el conjunto factible viene dado aquí por una desigualdad estricta. En cambio, el problema

$$\begin{array}{rl} \min\limits_{x} & e^x \\ \text{s.a.} & x \geq 0 \end{array}$$

es un problema con una restricción de desigualdad. Si desde el punto factible $x = 0$ nos movemos hacia la izquierda, ahí encontraremos puntos no factibles.

Antes hemos visto lo que en ocasiones recibe el nombre de *regla de Fermat*, que dice que, en un problema univariante de optimización sin restricciones, si la función objetivo es diferenciable, entonces la derivada debe valer cero en cualquier maximizador o cualquier minimizador. Lo enunciamos aquí formalmente para que quede constancia de ello.

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.

Observad que la regla de Fermat no proporciona un procedimiento para calcular automáticamente todos los maximizadores o minimizadores (locales) de una función, sino que sólo nos dice que podemos descartar como tales todos aquellos puntos en los que alguna derivada parcial sea diferente de cero. En general, necesitaremos criterios adicionales para seleccionar entre todos aquellos puntos que satisfacen esta primera criba, es decir, entre todos los puntos críticos de la función.

Pierre Simon de Fermat...

... (Beaumont-de-Lomagne 1601-Castus 1665) ideó su *Methodus ad Disquirendam Maximan et Miniman* en el año 1629.

Llamamos **condiciones de primer orden** en un problema de optimización sin restricciones al sistema de ecuaciones que describe los puntos críticos de la función objetivo, es decir, aquel que resulta de igualar a cero todas las derivadas parciales de la función.

Nuestro objetivo a continuación es proporcionar herramientas de análisis de los puntos críticos. En primer lugar veremos que, a partir de las condiciones de segundo orden, podremos saber en algunos casos si un punto crítico es un maximizador local, un minimizador local o un punto de silla. Después veremos que el análisis de la concavidad o convexidad de la función objetivo nos sirve para dar condiciones que garanticen que un punto crítico es un maximizador o un minimizador global.

2.2. Condiciones de segundo orden

En muchos casos, aunque no siempre, mediante el estudio de las derivadas parciales de segundo orden podemos determinar si la función crece o decrece alrededor de un punto crítico y, por lo tanto, si se trata de un **minimizador** o de un **maximizador**.

Lo primero que haremos es un pequeño repaso de las propiedades de las formas cuadráticas, ya que formularemos las condiciones de segundo orden estudiando formas cuadráticas compuestas a partir de las derivadas parciales.

2.2.1. Formas cuadráticas

En el curso de álgebra lineal ya habéis visto las definiciones de formas cuadráticas y su clasificación. Aquí repasamos las definiciones y fijamos la terminología que usaremos más adelante.

Sea

$$c = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix}$$

una matriz simétrica, es decir, que para todo $i \in \{1,2...n\}$ y para todo $j \in \{1,2,...,n\}$, se cumple $c_{ij} = c_{ji}$. Entonces podemos definir una función Q: $\mathrm{R}^n \rightarrow \mathrm{R}$ haciendo:

$$\forall \mathrm{v} \in \mathrm{R}^n, \quad Q(\mathrm{v}) = v' \ C \ v = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \ v_i \ v_j,$$

donde v' denota el vector v transpuesto. Diremos que Q es una **forma cuadrática**.

Ejemplo 2.5. Observad la relación entre la siguiente matriz C y la correspondiente forma cuadrática Q

$$c = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 4 \\ 3 & 4 & 7 \end{pmatrix}$$

$$Q(x,y,z) = x^2 + 4xy + 6xz + 5y^2 + 8yz + 7z^2.$$

Las filas 1, 2 y 3 de la matriz corresponden a las variables x, y y z, respectivamente. De forma análoga, las columnas 1, 2 y 3 también corresponden a las variables x, y y z, respectivamente. Dado que multiplicamos cada término dos veces (primero multiplicamos por el vector (x,y,x) y después premultiplicamos por el mismo vector transpuesto), la suma de los exponentes de las variables que aparecen en cada uno de los sumandos es 2. El coeficiente de x^2 es el término que corresponde a los índices (1, 1) de la matriz, en este caso el número 1. El coeficiente de y^2 es el término que corresponde a los índices (2, 2), en este caso el número 5. Finalmente, el coeficiente de z^2 es el término que corresponde a los índices (3, 3), en este caso el número 7.

El coeficiente del sumando que contiene el producto xy es el resultado de sumar el término (2,1) con el término (1,2), en este caso $2 + 2 = 4$. El coeficiente del sumando que contiene el producto xz es el resultado de sumar el término (3,1) con el término (1,3), en este caso $3 + 3 = 6$. Finalmente, el coeficiente del sumando que contiene el producto yz es el resultado de sumar el término (3,2) con el término (2,3), en este caso $4 + 4 = 8$.

La clasificación de formas cuadráticas que nosotros usaremos es la siguiente.

Sea Q: $\mathrm{R}^n \rightarrow \mathrm{R}$ una forma cuadrática.

1) Decimos que Q es **definida positiva** si, para todo $v \in \mathrm{R}^n$ tal que $v \neq 0$, se cumple $Q(v) > 0$.
2) Decimos que Q es **definida negativa** si, para todo $v \in \mathrm{R}^n$ tal que $v \neq 0$, se cumple $Q(v) < 0$.
3) Decimos que Q es **semidefinida positiva** si, para todo $v \in \mathrm{R}^n$ se cumple $Q(v) \geq 0$.
4) Decimos que Q es **semidefinida negativa** si, para todo $v \in \mathrm{R}^n$ se cumple $Q(v) \leq 0$.
5) Decimos que Q es **indefinida** si, existen dos vectores v y w de R^n tales que $Q(v) < 0 < Q\,(w)$.

Con frecuencia, una forma cuadrática se define como semidefinida si, además de la condición que hemos escrito arriba, hay algún vector $v \neq 0$ de modo que $Q(v) = 0$. Nosotros aquí no lo hacemos así. De acuerdo con la definición que acabamos de dar, toda forma definida positiva también es semidefinida positiva, mientras que una forma indefinida es aquella que no es semidefinida positiva ni semidefinida negativa. Utilizamos esta definición porque nos facilita mucho más la presentación de todos los resultados basados en formas cuadráticas que veremos más adelante. Cuando queramos hacer referencia a una forma cuadrática que es semidefinida positiva (o negativa) aludiremos a la misma diciendo que la forma cuadrática es sólo semidefinida.

Es importante tener en cuenta que la expresión $v \neq 0$ que aparece en la definición anterior quiere decir que si $v = (v_1, v^2 {}_{\dots} v^n)$, entonces hay al menos un índice i, de modo que el número v_i es diferente de cero.

En la práctica, trataremos con la matriz C más que con la correspondiente forma cuadrática Q. Por abuso de notación, diremos también que la *matriz C* es definida positiva, indefinida, etc.

Las condiciones que aparecen a continuación permiten verificar las definiciones anteriores trabajando sólo con la matriz.

Criterio de los valores propios

Sea $Q : \mathrm{R}^n \to \mathrm{R}$ una forma cuadrática, definida a partir de la matriz simétrica C como en la ecuación 1. Sean $\{\lambda_1, \lambda_2 \ldots \lambda_n\}$ los valores propios de la matriz C, entonces tenemos:

1) Q es definida positiva si, y sólo si, para todo i, $\lambda_i > 0$.
2) Q es definida negativa si, y sólo si, para todo i, $\lambda_i < 0$.
3) Q es semidefinida positiva si, y sólo si, para todo i, $\lambda_i \geq 0$.
4) Q es semidefinida negativa si, y sólo si, para todo i, $\lambda_i \leq 0$.
5) Q es indefinida si, y sólo si, hay índices i y j tales que $\lambda_i < 0 < \lambda_j$.

El teorema da una caracterización exacta de la categoría a la que pertenece la forma cuadrática Q a partir de los valores propios de la matriz C.

Si no tenemos un ordenador cerca, con frecuencia podemos ahorrarnos trabajo si usamos el teorema siguiente, que proporciona sólo una caracterización parcial, basada en los menores preferentes de la matriz C. Los menores preferentes son los determinantes de las sucesivas submatrices cuadradas que vamos formando partiendo del elemento superior izquierdo, hasta llegar a la matriz entera.

Criterio de los menores preferentes

Sea $Q : \mathrm{R}^n \to \mathrm{R}$ una forma cuadrática, definida a partir de la matriz C como en la ecuación 1. Sean $\{\Delta_1 \Delta_{2 \ldots} \Delta_n\}$ los menores preferentes de la matriz C. Entonces tenemos:

1) Q es definida positiva si, y sólo si, para todo i, $\Delta_i > 0$.
2) Q es definida negativa si, y sólo si, para todo i, $(-1)^i \Delta_i > 0$, es decir, si los menores preferentes de orden impar son negativos y los de orden par son positivos.

Dos observaciones:

- El criterio de los menores preferentes sólo es decisorio cuando la forma cuadrática es definida positiva o negativa. No permite distinguir formas cuadráticas indefinidas de otras que sólo son semidefinidas.

- La forma más sencilla de recordar las condiciones para que Q sea definida negativa es considerar el caso en que C es una matriz diagonal (y, por lo tanto, los elementos de la diagonal principal son sus valores propios); entonces tenemos: $D_1 = \lambda_1 < 0$, $D_2 = \lambda_1\lambda_2 > 0$, etc.

Ejemplo 2.6. Sea $Q(x,y) = 3x^2 + y^2$ y, por tanto,

$$c = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

es una matriz diagonal, que tiene por valores propios $\lambda_1 = 3 > 0$ y $\lambda_2 = 1 > 0$. El criterio de los valores propios implica que Q es definida positiva.

Alternativamente, en este caso es fácil analizar Q directamente a partir de la definición de forma cuadrática definida positiva. Observamos de entrada que $\forall(x,y)$, $3x^2 + y^2 \geq 0$, porque los cuadrados de números reales siempre son no negativos. Esto quiere decir que, como mínimo, Q es semidefinida positiva. Por otro lado, dado que el cuadrado de un número real es cero sólo cuando este número es cero, resulta sencillo darse cuenta de que Q sólo puede ser cero cuando tanto x como y son iguales a cero. Esto es precisamente lo que se requiere en la definición de forma cuadrática definida positiva.

Ejemplo 2.7. Sea $Q(x, y) = x^2 + xy + y^2$, de donde

$$c = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$$

Los menores preferentes son aquí $\Delta_1 = 1 > 0$ y $\Delta_2 = \frac{3}{4} > 0$, de modo que el criterio de los menores preferentes garantiza que Q también es definida positiva. Calcular los valores propios nos habría costado bastante más trabajo.

Si queremos analizar Q directamente, la podemos escribir como $Q(x,y) = \left(\frac{x}{2} + y\right)^2 + \frac{3}{4}x^2$, de donde resulta sencillo ver que es estrictamente positiva siempre que x e y no sean cero simultáneamente.

Ejemplo 2.8. Sea $Q(x, y) = 2xy$, de donde

$$c = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

El criterio de los menores preferentes no es aplicable, ya que $D_1 = 0$, por lo que tenemos que tener en cuenta los valores propios. Los valores propios son $\lambda_1 = 1$ y $\lambda_2 = -1$, lo cual indica que Q es indefinida.

A partir de la propia definición no es nada difícil probar que Q es indefinida; por ejemplo: $Q(-1,1) = -2 < 0 < 2 = Q(1,1)$. ¡Así de sencillo es el asunto! Releed la definición y veréis que, para verificar que Q es indefinida, tenemos suficiente encontrando dos elementos cualesquiera en los que Q toma signos opuestos.

Aparte de los criterios que acabamos de mencionar, vosotros habéis visto una forma diferente de clasificar formas cuadráticas, que se basa en los cambios de signo del polinomio característico. El criterio de los menores preferentes tiene la ventaja de que no necesitamos ni siquiera encontrar el polinomio característico para aplicarlo, y el inconveniente de que no siempre nos dice cómo es la forma cuadrática.

Hay otro criterio,...

... el de los menores principales, que es una extensión del de los menores preferentes, y que proporciona una caracterización completa del signo de la forma cuadrática; nosotros no lo damos aquí, pero lo podéis encontrar en muchos libros que hablan de optimización o de álgebra lineal.

2.2.2. Formas cuadráticas y matrices hessianas

Dado $X \subset \mathsf{R}^n$, sea f: $X \to \mathsf{R}$ una función que tiene derivadas parciales de segundo orden continuas, y sea $v = (v_1, v^2 \ldots\ v^n)$ un punto dado de R^n. $D^2 f(v)$ denota la matriz (hessiana) formada por las derivadas parciales segundas de f, evaluadas en el punto v.

La extensión a R^n de las condiciones de segundo orden que se estudian en el cálculo univariante está basada en el estudio de la matriz hessiana.

Condiciones suficientes de segundo orden

Sea $X \subset \mathsf{R}^n$ un conjunto abierto, y sea $f : X \to \mathsf{R}$ una función que tiene derivadas parciales de segundo orden continuas. Supongamos que $v \in \mathsf{R}^n$ es un punto crítico, es decir, que para cada i, $1 \le i \le n$, $D_i f(v) = 0$. Tenemos que:

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

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

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

Recordemos que...

... un punto x^* es un **maximizador** local de la función f, si hay un cierto entorno de x^* en el que se cumple que $f(x^*) \ge f(x)$, para cualquier punto x del entorno. Si la igualdad es estricta cuando $x \neq x^*$, decimos que x^* es un **maximizador** local estricto. El concepto de **minimizador** local estricto se define de forma similar.

Una observación importante: ninguna de las implicaciones del teorema anterior es cierta si la ponemos al revés. A continuación veremos algunos ejemplos de ello:

Ejemplo 2.9. Sea $f(x,y) = x^2 + y^2$. El único punto crítico es (0,0), y la matriz hessiana evaluada en este punto es

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

Dado que ésta es un matriz diagonal, los valores propios son $\lambda_1 = 2 > 0$ y $\lambda_2 = 2 > 0$ y, por lo tanto, es definida positiva. El teorema anterior implica que (0, 0) es un minimizador local estricto.

Ejemplo 2.10. Sea $f(x,y) = 1 - x^2 - y^2$. El único punto crítico es (0, 0), y la matriz hessiana evaluada en este punto es

$$D^2f(0,0) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

Dado que ésta es una matriz diagonal, los valores propios son $\lambda_1 = -2 < 0$ y $\lambda_2 = -2 < 0$ y, por lo tanto, es definida negativa. El teorema anterior implica que (0, 0) es un maximizador local estricto.

Ejemplo 2.11. Sea $f(x,y) = x^2 - y^2$. El único punto crítico es (0, 0) y la matriz hessiana evaluada en este punto es

$$D^2f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

Dado que ésta es una matriz diagonal, los valores propios son $\lambda_1 = 2 > 0$ y $\lambda_2 = -2 < 0$ y, por lo tanto, es indefinida. El teorema anterior implica que (0, 0) es un punto de silla.

Ejemplo 2.12. Sea $f(x, y) = x^4 + y^2$. El único punto crítico es (0, 0) y la matriz hessiana evaluada en este punto es

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

Al tratarse de una matriz diagonal, los valores propios son $\lambda_1 = 0$ y $\lambda_2 = 2 > 0$ y, por lo tanto, sólo es semidefinida positiva. El teorema anterior no es aplicable. Sin embargo, es fácil apreciar que, para todo $(x, y) \in \mathrm{R}^2$, $f(x,y) = x^4 + y^2 \geq 0 = f(0,0)$, con igualdad sólo cuando $(x,y) = (0, 0)$. Esto quiere decir que (0, 0) es un minimizador global (y, por lo tanto, local) estricto.

Ejemplo 2.13. Sea $f(x, y) = 1 - x^4 - y^2$. El único punto crítico es (0, 0), y la matriz hessiana evaluada en este punto es

$$D^2f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}$$

Dado que se trata de una matriz diagonal, los valores propios son $\lambda_1 = 0$ y $\lambda_2 = -2 > 0$ y, por lo tanto, sólo es semidefinida negativa. El teorema anterior no es aplicable. Sin embargo, es fácil ver que, para cada $(x, y) \in \mathrm{R}^2$, $f(x,y) = 1 - x^4 - y^2 \leq 1 = f(0, 0)$, con igualdad sólo cuando $(x, y) = (0, 0)$. Esto quiere decir que (0, 0) es un maximizador global (y, por lo tanto, local), estricto.

Ejemplo 2.14. Sea $f(x,y) = y^2 - x^4$. El único punto crítico es (0, 0) y la matriz hessiana evaluada en este punto es

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

Dado que ésta es una matriz diagonal, los valores propios son $\lambda_1 = 0$ y $\lambda_2 = 2 > 0$ y, por lo tanto, sólo es semidefinida positiva. El teorema anterior es aplicable. Sin embargo, para todo $x \neq 0$ e $y \neq 0$, tan cerca de 0 como queramos, se cumple $f(x,0) = -x^4 < 0 < y^2 = f(0,y)$, mientras que $f(0,0) = 0$. Esto quiere decir que (0, 0) es un punto de silla, aunque la matriz no sea indefinida.

Ejemplo 2.15. Sea $f(x,y) = x^4 - y^2$. El único punto crítico es (0, 0) y la matriz hessiana evaluada en este punto es

$$D^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}$$

Dado que ésta es una matriz diagonal, los valores propios son $\lambda_1 = 0$ y $\lambda_2 = -2 < 0$ y, por lo tanto, sólo es semidefinida negativa. El teorema anterior no es aplicable. Sin embargo, para todo $x \neq 0$ e $y \neq 0$, tan cerca de 0 como queramos, se cumple $f(x,0) = x^4 > 0 > -y^2 = f(0,y)$, mientras que $f(0,0) = 0$. Esto quiere decir que (0,0) es un punto de silla, aunque la matriz no sea indefinida.

Por lo tanto, las condiciones de segundo orden pueden no ser decisorias para determinar la naturaleza de un punto crítico. En algunos casos en los que el anterior teorema no es aplicable, podremos descartar al menos alguna posibilidad si tenemos en cuenta el siguiente resultado.

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 continuas. Entonces:

Si $v \in \mathbb{R}^n$ es un minimizador local (no necesariamente estricto), entonces v es un punto crítico y la matriz hessiana $D^2 f(v)$ es semidefinida positiva.

Si $v \in \mathbb{R}^n$ es un maximizador local (no necesariamente estricto), entonces v es un punto crítico y la matriz hessiana $D^2 f(v)$ es semidefinida negativa.

Observación importante

Ninguna de las implicaciones del teorema anterior es cierta si la ponemos al revés, como hemos podido ver en los ejemplos 2.14 y 2.15.

Como hemos podido apreciar en los ejemplos, las condiciones de segundo orden no son siempre decisorias. Entonces la pregunta es: ¿qué debemos hacer cuando nos encontramos en este caso? Desgraciadamente, no hay un método sistemático de análisis; hay una serie de técnicas que funcionan mejor en unos casos que en otros. Lo que se hace en general es proceder a lo que llamamos un análisis local del comportamiento de la función alrededor del punto crítico.

Los ejemplos que hemos presentado eran tan sencillos que hemos tenido suficiente con observar la función directamente; en general, tendremos que realizar un análisis considerando diferentes trayectorias de aproximación a la función.

Ejemplo 2.16. Supongamos que queremos analizar el punto crítico (0,0) de la función

$$f(x, y) = x^4y^4 + x^3y^3.$$

Dado que todas las variables tienen exponentes elevados, está claro que la matriz hessiana tendrá cero en todos los componentes cuando la evaluemos en el punto (0,0) y, por lo tanto, las condiciones de segundo orden no son decisorias. En este caso nos bastará con considerar lo que sucede a lo largo de dos trayectorias lineales de aproximación. Cuando aproximamos (0,0) a lo largo de la trayectoria lineal $y = x$, tenemos que el valor de la función es $f(x, x) = x^8 + x^6$, por lo que esta función presenta un mínimo local a lo largo de la trayectoria considerada. Cuando aproximamos (0, 0) a lo largo de la trayectoria inicial $y = -x$, tenemos que el valor de la función es $f(x, -x) = x^8 + x^6 = -x^6(1 - x^2)$, por lo que esta función presenta un máximo local a lo largo de la trayectoria considerada (la función vale 0 cuando $x = 0$, y toma valores negativos cuando $0<|x|<1$). Esto prueba que (0,0) es un punto de silla de la función.

Un último ejemplo nos servirá para comprender que, en ocasiones, el análisis local de la función puede ser bastante complicado.

Ejemplo 2.17. Supongamos que queremos analizar la función $f(x,y) = x^4 - 3x^2y + y^2$. Su único punto crítico es el (0,0), pero la matriz hessiana evaluada en este punto sólo es semidefinida, por lo que las condiciones de segundo orden no son decisorias. Para analizar trayectorias de aproximación lineal al punto, empezamos observando que, si nos aproximamos por los ejes (efectuamos $x = 0$ o $y = 0$), entonces la función presenta un mínimo en (0,0). Para analizar cualquier trayectoria lineal, supongamos que $y = tx$, donde t es un cierto número real (fijo). Entonces tenemos que $f(x,tx) = x^4 - 3tx^3 + t^2x^2$, y las herramientas habituales de cálculo muestran que f presenta un mínimo local en $x = 0$, cualquiera que sea t.

Todo esto parece sugerir que el punto crítico (0,0) es un minimizador local, pero si nos detenemos a considerar la trayectoria *cuadrática* de aproximación $y = x^2$, podemos comprobar que $f(x,x^2) = -x^4 < 0 = f(0,0)$, siempre que $x \neq 0$; es decir, f presenta un máximo en nuestro punto crítico a lo largo de esta trayectoria cuadrática de aproximación. Esto muestra que (0,0) es en realidad un punto de silla.

3. Concavidad y convexidad de funciones. Criterios de globalidad

3.1. Motivación. Caso univariante

Queremos encontrar un criterio que nos permita afirmar que cualquier maximizador o minimizador local lo es también en sentido global. De este modo, está claro que las condiciones que debemos imponer sobre el conjunto factible y la función objetivo deben ser también de carácter global; no nos podemos limitar a decir cómo debe ser la función objetivo justo alrededor de nuestro candidato a maximizador o minimizador global, sino que debemos exigir una cierta condición que se satisfaga en cualquier punto de la función objetivo.

Para entender los conceptos de concavidad y convexidad, lo primero que tenemos que hacer es analizar el comportamiento de las funciones alrededor de los puntos donde tienen máximos y mínimos (no importa si son locales o globales).

Ejemplo 3.1. Máximo, mínimo y función derivada

Sea $f(x) = x^3e^{-x^2}$. Antes hemos visto, en el ejemplo 1.19, que esta función tiene un máximo y un mínimo (locales y globales al mismo tiempo), y un punto de inflexión en medio de los dos. Lo que queremos ver ahora es cómo se comporta la derivada alrededor del máximo y del mínimo. Dejaremos que Gnuplot efectúe los cálculos. Lo que debemos hacer es introducir en el programa la función y su derivada (que hemos tenido que encontrar nosotros previamente):

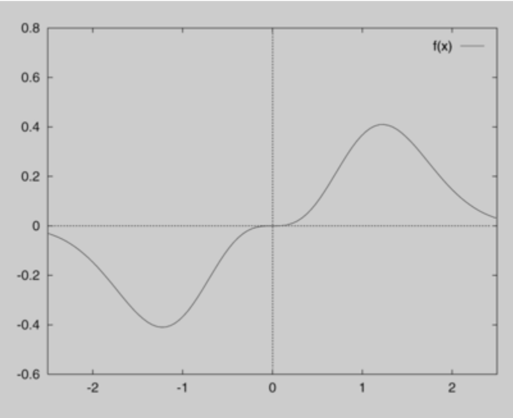
```
gnuplot> f(x) = x**3*exp(–x**2)
gnuplot> fp(x) = x**2**exp(–x**2)*(3–2*x**2)
```

Un poco de experimentación con los gráficos nos muestra cuáles son los valores de las variables más convenientes para la visualización:

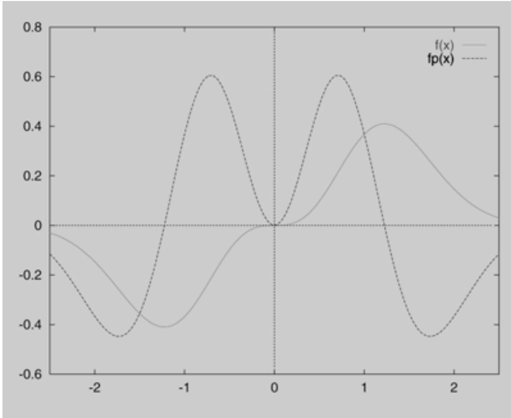
```
gnuplot> set xrange [–2.5:2.5]
gnuplot> set yrange [–0.6:0.8]
```

Primero confeccionamos una gráfica de la función *f* sola, para ver los puntos críticos, y a continuación la presentamos junto con su derivada:

```
gnuplot> plot f(x)
gnuplot> plot f(x), fp(x)
```



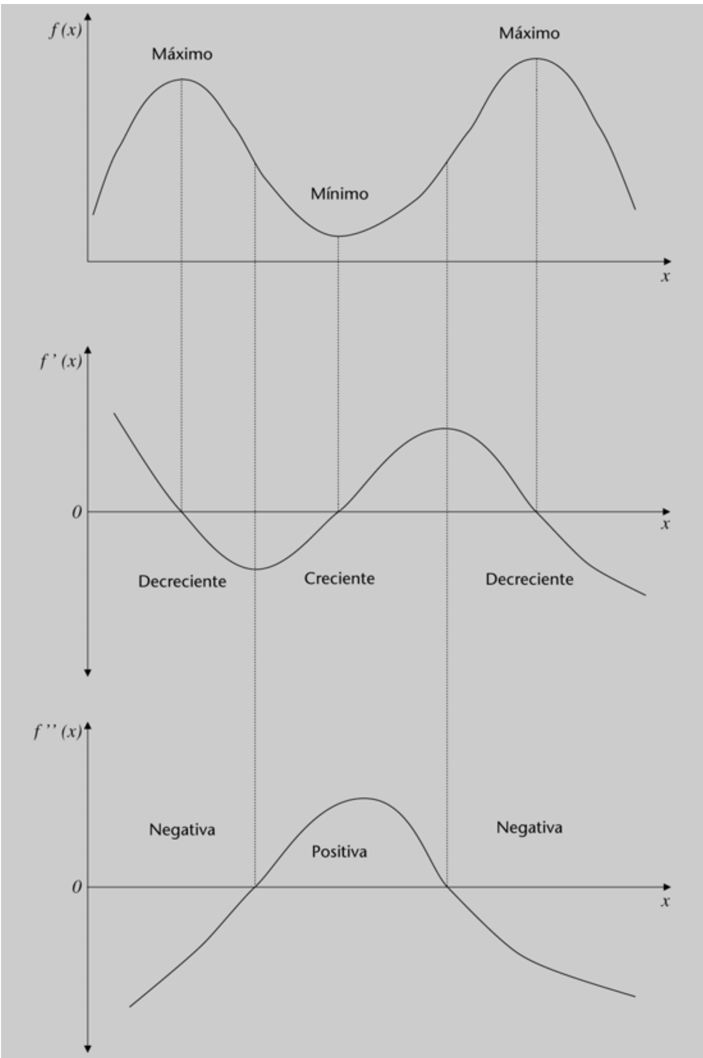
La función $x^3e^{-x^2}$ sola



Y con su derivada

La derivada es una función creciente alrededor del mínimo, y decreciente alrededor del máximo. Alrededor del punto de inflexión que aparece para $x = 0$, no es ninguna de las dos cosas.

En general, el comportamiento de una función y sus derivadas alrededor de los puntos máximos y los mínimos es parecido al que mostramos en la gráfica que tenemos a continuación.



Una función y sus derivadas

Un máximo local estricto

... es un punto en el que la función pasa de tener la derivada positiva a tenerla negativa, cuando nos vemos de izquierda a derecha. Es decir, alrededor del punto, la derivada es una función decreciente. Del mismo modo, alrededor de un mínimo local, la derivada es una función en la positividad o la negatividad de la derivada segunda.

Alrededor de un máximo (local), la derivada es una función decreciente y, por lo tanto, la derivada segunda es una función negativa. Si exigimos que la derivada sea una función decreciente en todas sus partes, entonces la función se comportará globalmente al igual que lo hace alrededor del máximo del que hemos partido; es decir, no podrá haber ningún otro máximo.

Una función cuya derivada es decreciente globalmente es una función cóncava. Más adelante veremos que los maximizadores locales de una función cóncava son siempre globales.

Del mismo modo, una función convexa es aquélla cuya derivada es una función creciente. Los minimizadores locales de una función convexa son siempre globales.

Consideremos una función cóncava. Fijemos un cierto punto a y seleccionemos otro punto x cualquiera. En este caso, sabemos por el teorema del valor medio que hay un punto c entre a y x tal que:

$$f'(c) = \frac{f(x) - f(a)}{x - a}$$

Si suponemos que $a < x$, entonces tenemos que $a < c$, y como la función es cóncava:

$$a < c \Rightarrow f'(a) \geq f'(c) \Rightarrow f'(a) \geq \frac{f(x) - f(a)}{x - a}$$

Multiplicando por $x–a$ (que es un número positivo), obtenemos:

$$f(a) + f'(a)(x–a) \geq f(x).$$

Por otro lado, si $a > x$, entonces $a > c$ y tendríamos:

$$a > c \Rightarrow f'(a) \leq f'(c) \Rightarrow f'(a) \leq \frac{f(x) - f(a)}{x - a}$$

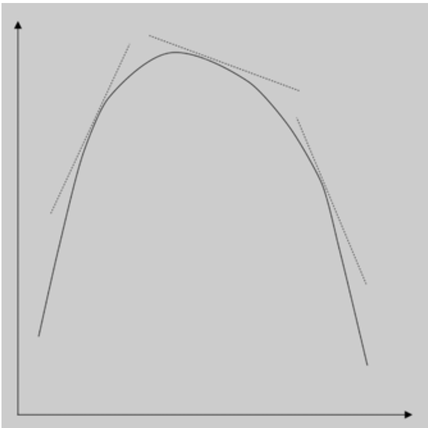
Pero ahora, al multiplicarse por $x - a$, tenemos que invertir la desigualdad, ya que este número es negativo. Con esto, nos vuelve a quedar

$$f(a) + f'(a)(x - a) \geq f(x).$$

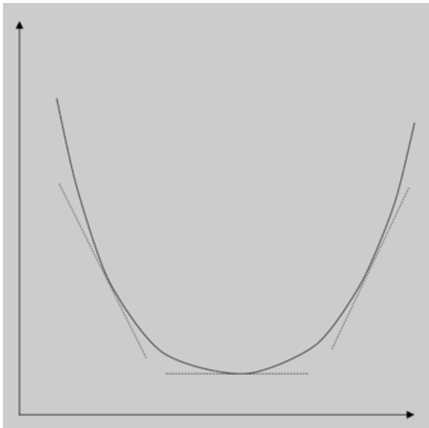
Resumiendo, cualesquiera que sean x y a, siempre se cumple la desigualdad anterior. Sin embargo, el término de la izquierda de la desigualdad es la expresión de la recta tangente a la función f en el punto a. Por lo tanto, la desigualdad nos está indicando que, si f es una función cóncava, la recta tangente siempre presenta un valor superior a la función. De hecho, esta comparación entre la función y sus tangentes no requiere ni siquiera que la función sea diferenciable; nosotros la tomaremos como definición provisional de concavidad.

Intuitivamente, una función cóncava es aquélla por la que las tangentes siempre tienen un valor superior o igual al de la función. Una función convexa es aquélla por la que las tangentes siempre tienen un valor inferior o igual al de la función.

Los conceptos que aparecen en la definición son puramente geométricos: el grafo de cualquier tangente en una función cóncava siempre queda por encima del grafo de la función. De modo análogo, el grafo de cualquier tangente siempre queda en una función convexa por debajo del grafo de la función.

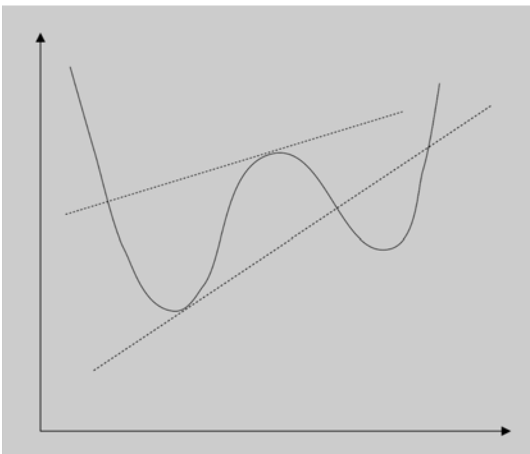


Una función cóncava



Una función convexa

Una función que no es cóncava ni convexa necesariamente debe ser cortada por alguna de sus tangentes.



Ni concavidad ni convexidad

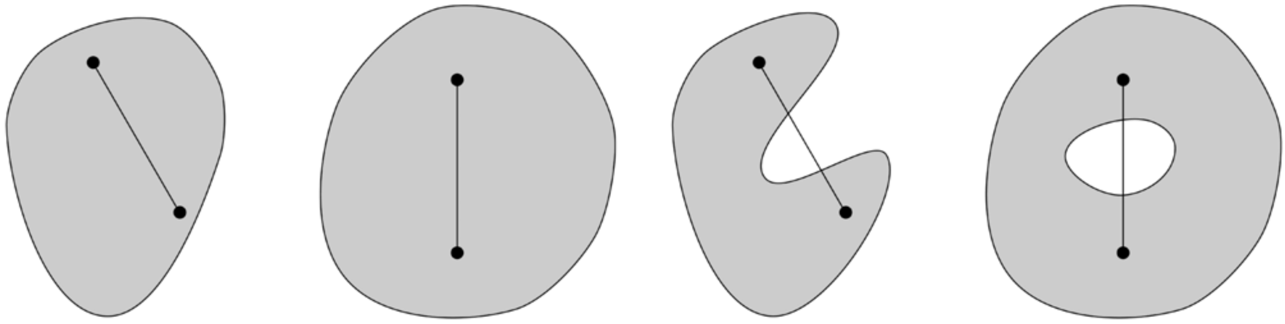
Alguna tangente...

... no está ni totalmente por debajo del gráfico ni totalmente por encima y, por lo tanto, lo corta.

3.2. Extensión a múltiples variables

Como primer requerimiento en las definiciones de funciones cóncavas y convexas impondremos que el dominio sobre el que las funciones están definidas (el conjunto factible) sea un conjunto convexo. Recordaremos su definición.

> Sea X un subconjunto de R^n. Decimos que X es un **conjunto convexo** si el segmento que une dos puntos cualesquiera del conjunto está enteramente contenido dentro de éste.



Conjuntos convexos

Conjuntos no convexos

El requerimiento de que el dominio de definición sea un conjunto convexo es muy importante a causa de todos los resultados que relacionan funciones cóncavas y convexas con optimización.

Se da una desafortunada coincidencia de las denominaciones que estamos usando, pero es muy importante tener presente lo siguiente:

Más adelante...

... veremos un ejemplo de cómo las cosas fallan cuando el dominio no es convexo.

> Una cuestión terminológica muy importante es que no se deben confundir los **conjuntos convexos** con las **funciones convexas** o las **funciones cóncavas**.

Si queremos comprobar algebraicamente si un conjunto es o no convexo, tendremos que ser capaces de describir el segmento que une dos puntos dados. La forma de hacerlo es la siguiente: dados dos puntos x' y x'', un punto x''' está sobre el segmento que los une si, y sólo si, hay un número $\lambda \in (0, 1)$, de modo que $x''' = \lambda x' + (1-\lambda)x''$.

Nosotros aquí tenemos suficiente con la noción intuitiva, por lo que no usaremos este tipo de formalización; así pues, no insistiremos en los aspectos algebraicos. Un resultado que sí usaremos es el siguiente.

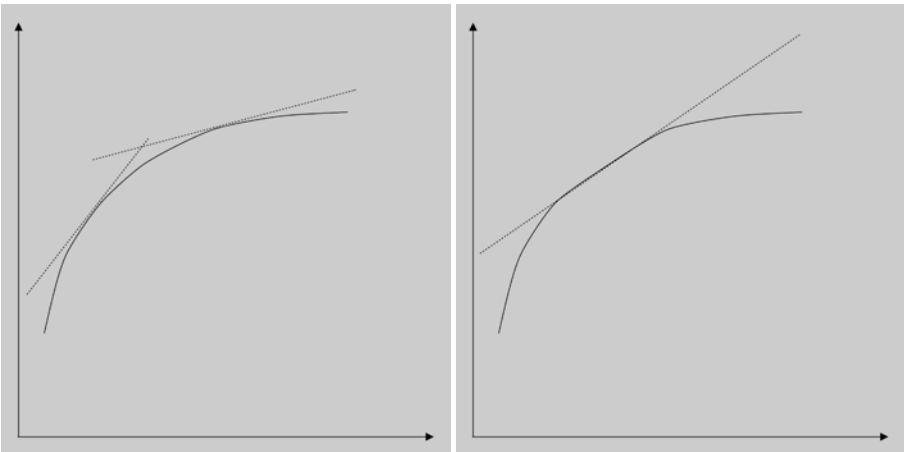
> La intersección de conjuntos convexos produce un conjunto que también es convexo.

La definición formal de función cóncava y función convexa es la misma que hemos dado antes, pero con el requerimiento de que el dominio de la función sea un conjunto convexo.

Sea X un subconjunto convexo de R^n, y $f : X \to R$ una función.

1) Decimos que *f* es una **función cóncava** si todas las tangentes quedan por encima del grafo de la función. La función es estrictamente cóncava si cada tangente tiene un único punto de contacto con el grafo.

2) Decimos que *f* es una **función convexa** si todas sus tangentes quedan por debajo del grafo de la función. La función es estrictamente convexa si cada tangente tiene un único punto de contacto con el grafo.



Función estrictamente cóncava

Función cóncava, no estrictamente

El grafo de una función que es cóncava pero no estrictamente tiene algún tramo plano.

Ahora supongamos que *f* es una función diferenciable. Denominaremos $\nabla f(x)$ al vector gradiente de la función *f* evaluado en el punto *x*, es decir, el vector compuesto por las derivadas parciales de la función *f* evaluada en el punto *x*. Recordemos que un punto "·" designa el producto escalar entre vectores. La tangente al grafo de *f* en el punto $(\bar{x}, f(\bar{x}))$ es el grafo de la función:

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

Por lo tanto, la condición geométrica que hemos dicho que satisface toda función cóncava es que, para todo $\bar{x} \in X$ y todo $x \in X$, se cumple que

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

Lo que aparece a continuación no es más que una forma de reescribir las definiciones para funciones diferenciables. Observamos que sólo se trata de la extensión a n dimensiones de lo que ya habíamos visto anteriormente.

Proposición 3. Sea $X \subset \mathsf{R}^n$ un conjunto convexo y f: $X \to \mathsf{R}$ una función diferenciable.

- f es una función cóncava si, y sólo si, dados dos puntos cualesquiera x′ y x″, se cumple que

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

- f es una función estrictamente cóncava si, y sólo si, dados dos puntos diferentes x′ y x'', se cumple que

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

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

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

- f es una función estrictamente convexa si, y sólo si, dados dos puntos diferentes x′ y x″, se cumple que

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

Decimos que una función es **continuamente diferenciable** si sus derivadas parciales son funciones continuas. De forma análoga, una función es **continuamente diferenciable dos veces** si sus derivadas parciales segundas son funciones continuas.

Un resultado ligeramente menos general, pero más útil para determinar cuándo una función es cóncava o convexa, es el siguiente, que requiere que la función sea continuamente diferenciable dos veces. Este resultado se deriva de la desigualdad del teorema anterior cuando se tiene en cuenta la expansión de Taylor de segundo orden.

Teorema

Sea $X \subset \mathsf{R}^n$ un conjunto abierto convexo y f : X → R una función continuamente diferenciable dos veces.

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

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

Lo que acabamos de ver es una condición equivalente a concavidad o a convexidad. Lo que aparece a continuación es un resultado útil, porque garantiza que una función sea cóncava o convexa en sentido estricto.

Teorema

Sea $X \subset \mathbb{R}^n$ un conjunto abierto convexo y $f : X \to \mathbb{R}$ una función continuamente diferenciable dos veces con continuidad.

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

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

Debemos tener en cuenta que hay funciones estrictamente cóncavas y funciones estrictamente convexas, de modo que su matriz hessiana sólo es semidefinida en algunos puntos aislados.

Consideremos, por ejemplo, la función $f(x) = x^4$, que es estrictamente convexa aunque $f''(0) = 0$. De hecho, la relación entre funciones estrictamente convexas y hessianas definidas positivas es la misma que la relación entre funciones estrictamente crecientes y derivadas positivas. Esto no es ninguna casualidad: recordemos que la convexidad para funciones univariantes quiere decir que la derivada es una función creciente.

Ejercicio

3.1. Aplicad el test de la segunda derivada para determinar si las funciones siguientes son cóncavas, convexas, o ninguna de las dos cosas.

a) $f(x) = x^2$

b) $f(x) = x^3$

c) $f(x) = x^3$, cuando $x > 0$

d) $f(x,y) = x + y$

e) $f(x,y) = x^2 - 2xy + y^2$

f) $f(x,y) = x^2 - xy + y^2$

g) $f(x,y) = x^4 + y^2$

h) $f(x,y) = 2x + 4y - 5 - x^2 - y^2$

3.3. Concavidad, convexidad y optimización

De los teoremas siguientes se deriva la importancia que tienen los conceptos de concavidad y convexidad en los problemas de optimización.

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^* también es un maximizador global.

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^* también es un minimizador global.

Observaciones:

- En un problema de maximización preferiremos que la función objetivo sea cóncava, ya que de este modo sabremos que los máximos locales también lo son en sentido global.
- En un problema de minimización preferiremos que la función objetivo sea convexa, ya que de este modo sabremos que los mínimos locales también lo son en sentido global.
- Observad que el maximizador o el minimizador local del que se habla no tiene por qué ser un punto interior. De hecho, el único requerimiento sobre el dominio es que éste sea un conjunto convexo. Esto nos permitirá utilizar estos teoremas para derivar resultados de globalidad en todo tipo de problemas de optimización, aunque tengan restricciones.
- Si una función es estrictamente cóncava, el maximizador del que habla el primer teorema es único. Si una función es estrictamente convexa, el minimizador del que habla el segundo teorema es único.

Ejemplo 3.2. Sea $X = \mathbb{R}_{++} = (0,\infty)$ y $f(x) = x \log(x)$. Tenemos que $f'(x) = 1 + \log(x)$ y $f''(x) = 1/x$. Dado que el dominio son los reales estrictamente positivos, tenemos que, para todo $x \in X$, se cumple $f''(x) = 1/x > 0$, por lo que podemos afirmar que f es una función convexa.

Los puntos críticos solucionan $f'(x) = 1 + log(x) = 0$, es decir, el único punto crítico es $x = 1/e$. Dado que $f''(1/e)>0$, sabemos que $1/e$ es un minimizador local, y el teorema anterior nos asegura que también es un minimizador global.

Ejemplo 3.3. Sea $X = \mathbb{R}^2$ *y* $f(x,y) = 1-x^2-y^2$. Las derivadas parciales *son* $D_1f(x,y) = -2x$ *y* $D^2f(x,y) = -2y$, y la matriz hessiana

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

Por lo tanto, la matriz hessiana es definida negativa para todo (x,y), y podemos concluir que f es una función cóncava.

Es sencillo darse cuenta de que el único punto crítico es el (0,0) y que éste es un maximizador local, porque $D^2f(0,0)$ es definida negativa. El teorema que hemos visto anteriormente nos asegura que (0,0) también es un maximizador global.

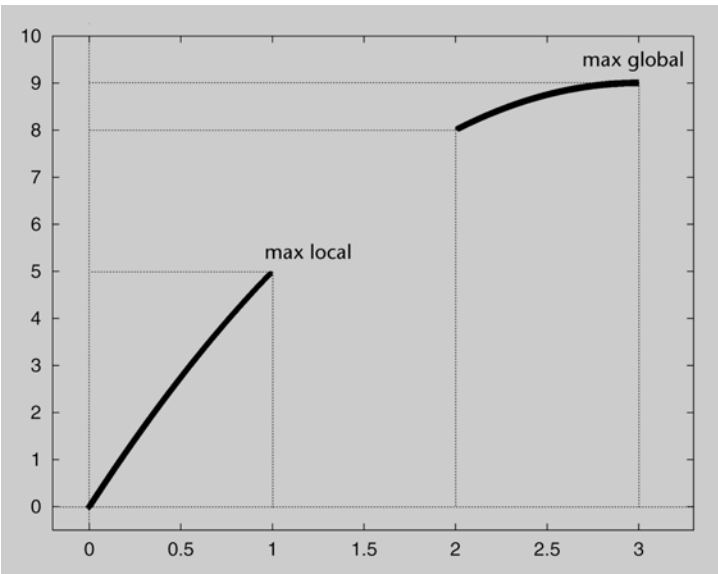
Debemos tener presente que resulta esencial que el dominio de la función sea un conjunto convexo.

Ejemplo 3.4. El dominio no es convexo

Sea $f(x) = 6x - x^2$. Es sencillo ver, aplicando el criterio de la derivada segunda, que esta función es estrictamente cóncava. Supongamos que queremos resolver el problema de maximizar f sobre el conjunto $X = [0,1] \cup [2,3]$. Dado que X es la unión de dos intervalos que no se tocan, se trata de un conjunto no convexo.

Dentro de este dominio de definición podemos comprobar rápidamente que el punto $x^* = 1$ es un maximizador local; esto se debe al hecho de que los únicos puntos con los que podemos compararlo localmente son los del intervalo [0,1]. También resulta sencillo apreciar que $x^* = 1$ *no* es un maximizador global; de hecho, el único maximizador global es $x^{**} = 3$.

¿Cuál es el problema de este ejemplo? Debéis notar que si miramos la definición formal de concavidad, observaremos que un requerimiento para afirmar que una función es cóncava consiste en que su dominio sea un conjunto convexo. Es decir, que aunque la segunda derivada de f sea estrictamente negativa, el hecho de que su dominio no sea convexo implica que no podemos decir que la función sea cóncava, como hemos hecho alegremente al empezar el ejemplo.



Local pero no global

El punto $x^* = 1$...

... es un maximizador local, pero el maximizador global es $x^{**} = 3$.

Si obervamos la caracterización de funciones cóncavas y de funciones convexas que son diferenciables, podemos deducir inmediatamente otro criterio que también resulta muy útil.

Teorema

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

1) Si f es una función cóncava, entonces cualquier punto en el que todas las derivadas parciales son cero es un maximizador global.

2) Si f es una función convexa, entonces cualquier punto en el que todas las derivadas parciales son cero es un minimizador global.

Observemos que...

... este último resultado se aplica incluso en problemas donde los criterios locales por problemas de optimización sin restricciones no nos permiten decidir si un punto crítico es, localmente, un maximizador, un minimizador o un punto de silla.

Ejemplo 3.5. Sea $f(x,y) = x^4 + y^4$. Es fácil apreciar que el único punto donde se anulan las derivadas parciales es el (0,0). La matriz hessiana es

$$D^2 f(x, y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 12y^2 \end{pmatrix}$$

y es semidefinida para todo $(x,y) \in \mathrm{R}^2$. Esto significa que la función es convexa y que, por lo tanto, el único punto crítico (0,0) es un minimizador global, aunque las condiciones de segundo orden para problemas de optimización sin restricciones ni siquiera nos permiten decidir si se trata de un minimizador local, porque la matriz hessiana en (0,0) es una matriz de ceros.

3.4. Solucionario

3.1.

a) $f(x) = x^2 \to f''(x) = 2 > 0$. Estrictamente convexa.

b) $f(x) = x^3 \to f''(x) = 3x$. Positiva o negativa. Ni cóncava ni convexa.

c) $f(x) = x^3 \to f''(x) = 3x > 0$. Estrictamente convexa.

d) $f(x,y) = x + y$. La matriz hessiana es

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

que es tanto semidefinida positiva como semidefinida negativa. Cóncava y convexa.

e) $f(x,y) = x^2 - 2xy + y^2$. La matriz hessiana es

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

es semidefinida positiva. Convexa.

f) $f(x,y) = x^2 - xy + y^2$. La matriz hessiana es

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

es definida positiva. Estrictamente convexa.

g) $f(x,y) = x^4 + y^2$. La matriz hessiana es

$$D^2 f(x,y) = \begin{pmatrix} 12x^2 & 0 \\ 0 & 2 \end{pmatrix}$$

que es definida positiva (excepto en el origen). Estrictamente convexa.

h) $f(x,y) = 2x + 4y - 5x^2 - y^2$. La matriz hessiana es

$$D^2 f(x,y) = \begin{pmatrix} -10 & 0 \\ 0 & -2 \end{pmatrix}$$

que es definida negativa. Estrictamente cóncava.

4. Optimización con restricciones de igualdad

4.1. Condiciones de primer orden: el método de los multiplicadores de Lagrange

Un problema de optimización con restricciones de igualdad es aquél en el que el conjunto factible está formado por todos los puntos que solucionan un determinado sistema de ecuaciones. Por ejemplo,

$$\begin{array}{ll} \max\limits_{(x,y)} & xy \\ \text{s.a.} & x+y=1. \end{array}$$

En este problema, al ser la restricción una ecuación lineal, podemos expresar una de las variables en términos de la otra y solucionar el problema (sin restricciones) que resulta después de sustituir esta variable en la función objetivo. Por ejemplo, dado que cualquier punto factible satisface $y = 1-x$, resulta que debemos solucionar el problema sin restricciones

$$\begin{array}{ll} \max\limits_{x} & x(1-x) \\ \text{s.a.} & x \in \mathsf{R}, \end{array}$$

que tiene por solución única y global $x^* = 1/2$ y, por lo tanto, $(x^*,y^*) = (1/2, 1/2)$ es el maximizador del problema original.

Cuando las restricciones estén definidas por ecuaciones no lineales, como por ejemplo en el problema

$$\begin{array}{ll} \max\limits_{(x,y)} & xy \\ \text{s.a.} & x^2+y^2=2, \end{array}$$

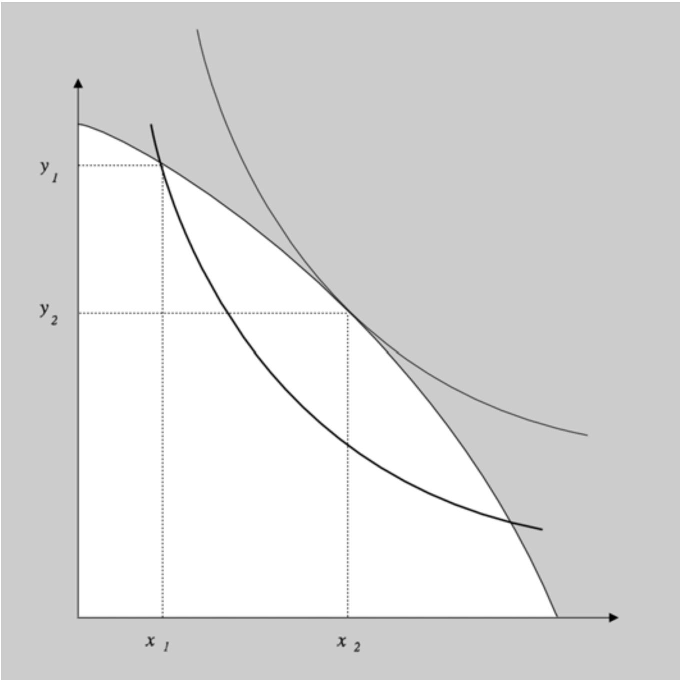
podemos encontrar las condiciones necesarias de primer orden siempre que el teorema de la función implícita sea aplicable a los maximizadores o los minimizadores. En estas notas no seguiremos este camino, sino que derivaremos las condiciones necesarias a partir de consideraciones geométricas.

La idea básica de esta derivación es sencilla. Recordad que, dada una función $f: \mathsf{R}^n \to \mathsf{R}$ y dado cualquier $v \in \mathsf{R}^n$, la curva de nivel que pasa por v separa los puntos donde la función toma valores superiores a $f(v)$ de aquellos puntos donde la función toma valores inferiores. Considerad ahora un problema de optimización con una restricción de igualdad. El conjunto factible está formado por todos aquellos puntos que satisfacen la restricción; con n variables, el conjunto factible consistirá en una cierta superficie $n - 1$ dimensional. Si en un punto factible la curva de nivel de la función objetivo corta la superficie factible, entonces este punto no puede ser maximizador ni minimizador, ya que hay puntos factibles en los dos lados de la curva de nivel.

Por lo tanto, una condición necesaria para que un punto sea un maximizador o un minimizador local será que la curva de nivel que pasa por este punto sea tangente a la superficie definida por el conjunto factible. Por ello decimos que esta condición necesaria es la condición de tangencia.

Fijaos en ello

En (x_1,y_1), la curva de nivel corta el conjunto factible. En (x_2,y_2) se da tangencia.



Tangencia

Volvemos al problema que habíamos formulado anteriormente, para ver la translación en términos analíticos de la condición de tangencia.

Definimos $f(x,y) = xy$, la función objetivo, y $g(x,y) = x^2 + y^2 - 2$, de modo que la restricción es $g(x,y) = 0$, la curva de nivel 0 de la función g. Un punto factible (x^*,y^*) satisface la condición de tangencia si en este punto las curvas de nivel de f y de g son tangentes. En el contexto que estamos considerando, en el que las curvas son definidas por funciones diferenciables, dos curvas son tangentes en un punto si hay una sola recta tangente a cada una de las curvas en este punto. Dado que la tangente a una curva de nivel es perpendicular al vector gradiente, la condición de tangencia equivale a decir que los vectores gradientes de f y de g en el punto (x^*,y^*) son colineales, es decir, que existe un cierto número real λ tal que

$$\nabla f(x^*,y^*) = \lambda \nabla g(x^*,y^*) \Rightarrow \begin{cases} \dfrac{\partial f}{\partial x} = \lambda \dfrac{\partial g}{\partial x} \\ \dfrac{\partial f}{\partial y} = \lambda \dfrac{\partial g}{\partial y} \end{cases}$$

Observad que ahora tenemos una variable adicional, l, que denominaremos **multiplicador de Lagrange** del problema. Si le añadimos la restricción, obtenemos un sistema con tres ecuaciones y tres incógnitas:

$$y = \lambda(2x)$$
$$x = \lambda(2y)$$
$$x^2 + y^2 = 2.$$

La solución de este sistema nos da cuatro candidatos a maximizadores o minimizadores locales: (1,1), (–1,–1), (1,–1) y (–1,1). Un análisis gráfico nos muestra sin demasiadas dificultades que los dos primeros son maximizadores globales y los dos últimos, minimizadores globales.

Consideremos ahora el caso general en el que participan más de una restricción. Observad que el número de restricciones independientes no puede exceder el número de variables; intuitivamente, cada restricción que añadimos reduce el problema en una dimensión y, por lo tanto, el problema no tendrá sentido si no quedan dimensiones libres o “grados de libertad”. Sin embargo, si el número de variables es igual que el de restricciones independientes, entonces el problema es trivial, pero todavía tiene sentido formularlo (esto nos servirá para cuando tratemos las restricciones de desigualdad).

Cuando se da más de una restricción, la condición de tangencia todavía es necesaria: si la curva de nivel de la función objetivo corta el conjunto factible, entonces habrá puntos factibles en los que la función objetivo toma valores mayores y otros en los que toma valores menores.

Lo que ahora requiere una explicación adicional es la expresión de la condición de tangencia en términos de los vectores gradientes de la función objetivo y de las restricciones. Supongamos que hay n variables y m restricciones, con $m < n$. Denominamos f la función objetivo y $g_1, g^2 \ldots g_m$, las funciones que definen las restricciones, tal como las hemos definido antes (la i–ésima restricción es la curva de nivel 0 de g_i). Las curvas de nivel de f tienen n–1 dimensiones dentro de R^n, porque están definidas por una ecuación. El conjunto factible, que está definido por m ecuaciones dentro de R^n, tiene dimensión n–m. La condición de tangencia requiere que el plano tangente (de dimensión n–1) en la curva de nivel de f contenga el plano tangente (de dimensión n–m) en el conjunto factible. Cuando lo miramos en términos de las perpendiculares de los vectores gradientes, las relaciones se invierten: si son linealmente independientes, los m gradientes de las restricciones forman un plano m dimensional; si el gradiente de f está contenido dentro de este plano, entonces el plano tangente en la curva de nivel de f contendrá el plano tangente en el conjunto factible. La condición que buscamos es, de este modo, que el gradiente de la función objetivo sea una combinación lineal de los gradientes de las restricciones.

$$\nabla f(v) = \lambda_1 \nabla_{g1}(v) + \lambda_2 \nabla_{g2}(v) + \ldots + \lambda_m \nabla g_m(v)$$

Junto con las m restricciones, esto nos proporciona un sistema con $n + m$ ecuaciones en el que las $n + m$ incógnitas son las n variables y los m multiplicadores de Lagrange. Observad que hay un multiplicador de Lagrange asociado a cada restricción.

Para garantizar la validez general de la condición de tangencia, necesitamos una condición técnica que nos asegure que los planos tangentes se pueden representar a partir de los vectores gradientes de las funciones respectivas. Esta condición consiste en que los gradientes de las restricciones son linealmente independientes. Más adelante estudiaremos con mayor detalle el significado de esta condición.

Regla de los multiplicadores de Lagrange

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

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

$$\begin{gathered} g_1(v_1,v_2,...,v_n) = 0 \\ g_2(v_1,v_2,...,v_n) = 0 \\ \dots\dots\dots\dots\dots\dots \\ g_m(v_1,v_2,...,v_n) = 0 \end{gathered}$$

Si los gradientes de g_1,g_1... gm evaluados en v^* son linealmente independientes, entonces hay números reales $\lambda_1\lambda_2$ ... λ_m, que cumplen

$$\nabla f(v^*) = \lambda_1 \nabla g_1(v^*) + \lambda_2 \nabla g_2(v^*) + \dots + \lambda_m \nabla g_m(v^*).$$

Observaciones:

- Las condiciones que hemos derivado son necesarias, pero no suficientes. Esto quiere decir que, en general, cualquier maximizador o minimizador las satisfará, pero también puede haber soluciones que no sean ninguna de las dos cosas.

- Dado que hemos aproximado las funciones mediante la diferenciación, los maximizadores o minimizadores en principio sólo lo serán localmente. Necesitaremos criterios adicionales para determinar su carácter global.

- Para que las condiciones se cumplan, será necesario que las tangentes estén bien definidas. Esto será cierto siempre que los gradientes de las restricciones sean linealmente independientes cuando los evaluemos en los maximizadores o los minimizadores.

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

Dadas las funciones f, g_1, $g_2 ... g_m$, tal como en el teorema anterior, definimos el lagrangiano o función de Lagrange como la función de $n + m$ argumentos:

$L(v_1,v_2...v_n, \lambda_1, \lambda_2... \lambda_m) =$

$f(v_1,v_2...vn)- \lambda_1 g_1(v_1,v_2...v_n)-...$

$...- \lambda_2 g_2(v_1,v_2...v_n)-...- \lambda_m g_m(v_1,v_2...v_n).$

Si interpretamos de forma intuitiva el lagrangiano, veremos que consiste en una función objetivo sustitutoria, que nos permite olvidarnos del hecho de que las variables tienen que obedecer las restricciones.

Para aplicar la regla de los multiplicadores de Lagrange, en primer lugar definimos el lagrangiano y después igualamos a cero las derivadas parciales de éste respecto a cada una de las variables originales. Finalmente, añadimos las restricciones para obtener un sistema con tantas ecuaciones como incógnitas.

Construimos el lagrangiano...

... añadiendo a la función objetivo original un término de penalización para cada restricción, de modo que los términos adicionales de penalización sólo tienen efecto cuando las respectivas restricciones no se cumplen.

Por ejemplo, en el caso de dos variables (x,y) y una restricción, tenemos $L(x,y, \lambda) = f(x,y)- \lambda g(x,y)$. Igualando a cero las derivadas parciales del lagrangiano obtenemos:

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0 \text{ y}$$
$$\frac{\partial L}{\partial y} = \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0,$$

que son exactamente las condiciones de tangencia.

Ejemplo 4.1. Queremos encontrar los maximizadores o los minimizadores de: $f(x,y,z) = x + y + 2z$ sobre la superficie definida por

$$x^2 + y^2 + z^2 = 2$$
$$x + y + z = 2$$

Definimos $g_1(x,y,z) = x^2 + y^2 + z^2 - 2$ y $g^2(x,y,z) = x + y + z - 2$. Por lo tanto, el lagrangiano es:

$$L(x,y,z, \lambda_1, \lambda_2) = x + y + 2z-\lambda_1(x^2 + y^2 + z^2 - 2)- \lambda_2(x + y + z - 2).$$

Para aplicar el método de los multiplicadores de Lagrange, igualamos a cero las derivadas parciales del lagrangiano respecto a x, y, z, y después añadimos las restricciones:

$$\begin{aligned}
\frac{\partial L}{\partial x} &= 1 - 2\lambda_1 x - \lambda_2 = 0 \\
\frac{\partial L}{\partial y} &= 1 - 2\lambda_1 y - \lambda_2 = 0 \\
\frac{\partial L}{\partial z} &= 2 - 2\lambda_1 z - \lambda_2 = 0 \\
x^2 + y^2 + z^2 &= 2 \\
x + y + z &= 2
\end{aligned}$$

La primera ecuación y la segunda implican:

$$\lambda_1(x-y) = 0.$$

Si hacemos $\lambda_1 = 0$ en la segunda ecuación y en la tercera, esto nos lleva a una contradicción; por lo tanto, podemos concluir que $x = y$. Sustituyéndolo en las dos restricciones, acabamos por obtener dos soluciones:

$$(x,y,z,\lambda_1,\lambda_2) = (1,1,0,-1/2,2)$$
$$(x,y,z,\lambda_1,\lambda_2) = (1/3,1/3,4/3,1/2,2/3).$$

Una interpretación geométrica (el conjunto factible resulta del hecho de cortar un casquete en diagonal a una esfera centrada en el origen de $\mathbb{R}^3$) nos permite ver que el primer punto es un minimizador global y el segundo, un maximizador global.

Ejemplo 4.2. Debéis de estar familiarizados con funciones de una sola variable, llamémosla $h(x)$, donde las soluciones de $h'(x) = 0$ no son maximizadores ni minimizadores, o bien sólo lo son localmente. Para ver que en los problemas con restricciones de igualdad nos podemos encontrar con extremos locales que no lo son globalmente, o con puntos críticos que no son maximizadores ni minimizadores, sólo es necesario que observemos que los dos problemas siguientes son equivalentes:

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

Nota

Os dejamos la tarea de probar que, si aplicamos el método de Lagrange en el 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 ni maximizador ni minimizador. Por otro lado, tomando $h(x) = 5x^2 - 3x^4 + \frac{1}{3}x^6$, encontraremos maximizadores y minimizadores que lo son localmente, pero no globalmente.

4.2. Regularidad

Hemos remarcado antes que, para aplicar el método de Lagrange, necesitamos que los gradientes de restricciones evaluados en el maximizador o el minimizador sean linealmente independientes. Cuando esta condición no se cumple, los gradientes de las restricciones no describen bien el plano tangente en el conjunto factible, y las cosas fallan: se pueden dar soluciones al problema de optimización que no aparezcan cuando aplicamos el método de Lagrange.

Ejemplo 4.3. Consideremos el problema

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

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

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

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

Observad 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 $\nabla g(0,0) = (0,0)$, linealmente dependiente (recordemos que el vector de ceros nunca forma parte de un conjunto de vectores linealmente independientes).

Geométricamente, dado que la restricción es el eje de las x, el plano tangente a la restricción en cualquier punto continúa siendo el eje de las x. Sin embargo, como el gradiente de g es el vector 0, el conjunto de vectores ortogonales en el gradiente de g no es el plano tangente, sino todo el espacio $\mathbb{R}^2$.

> Decimos que un punto factible es **regular** si los vectores gradientes de las restricciones son linealmente independientes cuando los evaluamos en este punto.

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

Observad 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 en los que falle la condición de regularidad o en qué casos dejará de hacerlo. Lo ilustraremos con un par de ejemplos.

Ejemplo 4.4. Consideremos un problema con dos variables y la restricción $g(x,y) = x^2 + y^3 = 0$. Dado que $\nabla g(x,y) = (2x, 3y^2)$, todos los puntos factibles son regulares, a excepción del punto (0,0). Dada una función f que tiene un maximizador o un minimizador local sobre esta superficie en (0,0), queremos saber si el método de Lagrange tendrá este punto como solución o no. Debéis notar que todo se reduce a verificar si se cumple la condición de tangencia $\nabla f(0,0) = \lambda\ \nabla g(0,0)$. Dado que $\nabla g(x,y) = (0,0)$, 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 Lagrange no funcionará, pero sí lo hará si $f(x,y) = x^2 + y^2$.

De este modo, en el caso de una sola restricción, el método de Lagrange detectará un maximizador no regular o un minimizador no regular sólo si todas las derivadas parciales en este punto son nulas. Cuando hay más de una restricción, también se producirán otros casos.

Ejemplo 4.5. Consideremos las dos restricciones

$$g_1(x,y,z) = (x-1)^2 + y^2 = 1$$

$$g_2(x,y,z) = (x+1)^2 + y^2 = 1.$$

Aquí tenemos $\nabla g_1(x,y,z) = (2(x-1),2y,0)$ y $\nabla g_2(x,y,x) = (2(x+1),2y,0)$.

Dado que el conjunto factible es

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

todos los puntos factibles satisfacen

$$\nabla g_1(x,y,z) = (-2,0,0)$$

$$\nabla g^2(x,y,z) = (2,0,0).$$

Es decir, no hay ningún punto factible que sea regular. La función $f(x,y,z) = x$ es constante sobre el conjunto factible y, de este modo, 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 en la curva de nivel de la función objetivo en el punto crítico contenga el eje de las z, que conforman el conjunto factible; sin embargo, la dependencia lineal de los gradientes de las restricciones hace que analíticamente sólo podamos detectar uno de los muchos planos tangentes que satisfacen esta propiedad, que es el plano y–z. El método de Lagrange funcionará sólo si el gradiente de f en el punto crítico es ortogonal al plano y–z.

4.3. Criterios de globalidad

En el caso con restricciones de igualdad, si queremos aplicar directamente el criterio de concavidad/convexidad, sólo lo podremos hacer cuando todas las restricciones sean **ecuaciones lineales**, porque es esencialmente el único caso (ved la nota al margen) en que el conjunto factible es convexo.

Cuando el conjunto factible sea convexo, el teorema local-global nos permite inferir el carácter global de los maximizadores o los minimizadores locales. Sin embargo, hay otra manera de aplicar los criterios de concavidad/convexidad, que es mediante un análisis del lagrangiano. A continuación veremos cómo analizando la curvatura del lagrangiano obtendremos un criterio de globalidad.

Nota

Exceptuando restricciones que tengan más de una representación, por ejemplo, la restricción lineal $y = 0$, que también puede ser escrita de forma no lineal como $y^3 = 0$. Para darnos cuenta de que esta representación es anómala sólo es necesario repasar el ejemplo en que este efecto hacía que el método de los multiplicadores de Lagrange no fuera aplicable.

Teorema

Sea (v^*,λ^*) un punto que soluciona las condiciones de primer orden del método de Lagrange, en el que usaremos la notación vectorial $v = (v_1,v_2...v_n)$ y $\lambda = (\lambda_1, \lambda_2... \lambda_m)$. Definimos la función $h(v) = L(v, \lambda^*)$; es decir, h es el lagrangiano cuando fijamos los multiplicadores en los valores λ^* y tomamos el vector v como variable. Entonces tenemos:

1) Si h es una función cóncava, v^* es un maximizador global del problema con restricciones de igualdad.
2) Si h es una función convexa, v^* es un minimizador global del problema con restricciones de igualdad.

Demostración del teorema:

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

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

es decir, todas las derivadas parciales de h son 0 cuando las evaluamos en el vector v^*. A causa de que h es una función cóncava, v^* es un maximizador global de h, es decir, $h(v^*) \geq h(v)$, para todo v del dominio (satisfaga o no las restricciones de igualdad). Tenemos que, para todo v:

$$h(v^*) \geq h(v) \Leftrightarrow f(v^*) - \sum_{i=j}^{m} \lambda_i^* \quad g_i(v^*) \geq f(v) - \sum_{i=1}^{m} \lambda_i^* \quad g_i(v).$$

Dado que v^* satisface las restricciones de igualdad, tenemos que, para cada $i = 1,2 \ldots m, g_i(v^*) = 0$, y esto implica que

$$f(v^*) \geq f(v) - \sum_{i=j}^{m} \lambda_i^* \quad g_i(v).$$

En particular, si v es un punto factible del problema con restricciones de igualdad, también se cumple que, para cada $i = 1, 2, \ldots m$, $g_i(v) = 0$, y eso implica que

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

Es decir, v^* es un maximizador global del problema con restricciones de igualdad. Una prueba paralela se utiliza en el caso de minimización.

Ejemplo 4.6. Sea $f(x,y) = x\,y, g(x,y) = x^2 + y^2 - 2$, y consideremos el problema de encontrar los máximos y mínimos de f sobre la superficie definida por $g(x,y) = x^2 + y^2 - 2 = 0$. El lagrangiano es $L(x,y,\lambda) = xy - \lambda(x^2 + y^2 - 2)$, 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, en las que (x, y, λ) toman los valores

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

Para analizar el carácter global de las soluciones, empecemos por la primera: $(x, y, \lambda) = (1, 1, 1/2)$. Para definir la función h de la que habla el teorema, fijamos $\lambda = 1/2$ y dejamos (x,y) como variables en el lagrangiano.

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

Las derivadas de primer orden de h son

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

En particular, como ya sabíamos por las condiciones de tangencia, se cumple

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

La matriz hessiana de h es

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

que podéis comprobar con facilidad que es semidefinida negativa. Dado que esta expresión es cierta para todo $(x,y) \in \mathbb{R}^2$, sabemos que h es una función cóncava sobre $\mathbb{R}^2$. Dado que 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) también es un maximizador global de la función $f(x,y) = xy$ sobre la superficie definida por $g(x,y) = x^2 + y^2 - 2 = 0$.

Un análisis similar...

... para cada uno de los otros puntos nos conduce a la conclusión de que tanto (1,1) como (–1,–1) son maximizadores globales, mientras que (1,–1) y (–1,1) son minimizadores globales.

Ejemplo 4.7. Recordemos el ejemplo que habíamos visto anteriormente, en el que $f(x,y) = -x^2 + y$, $g(x,y) = x^2 + y^2 - 1 = 0$. Las condiciones de primer orden nos dan cuatro puntos críticos:

$$(x^*,y^*,\lambda^*) = \left(0, 1, \frac{1}{2}\right)$$

$$(x^*,y^*,\lambda^*) = \left(0, -1, -\frac{1}{2}\right)$$

$$(x^*,y^*,\lambda^*) = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1\right)$$

$$(x^*,y^*,\lambda^*) = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1\right)$$

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

Consideremos el primer punto $(x^*.y^*,\lambda^*) = \left(0, 1, \frac{1}{2}\right)$. Sustituimos $\lambda = \frac{1}{2}$ en el lagrangiano para formar la función h que le corresponde: $h(x,y) = -x^2 + y - \frac{1}{2}(x^2 + y^2 - 1)$. La matriz hessiana de h es

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

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

Por otro lado, si tomamos el punto $(x^*.y^*,\lambda^*) = \left(0, -1, -\frac{1}{2}\right)$, y sustituimos $\lambda = -\frac{1}{2}$ en el lagrangiano, obtenemos $h(x,y) = -x^2 + y + \frac{1}{2}(x^2 + y^2 - 1)$. Ahora, la matriz hessiana de h es

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

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

Si finalmente consideramos el punto $\left(\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1\right)$ o el $\left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1\right)$, dado que tienen la misma $\lambda = -1$, les corresponde a los dos la misma función $h(x,y) = -x^2 + y + (x^2 + y^2 - 1)$. La matriz hessiana de h es

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

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

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

Corolario

Supongamos que, en un problema con restricciones de igualdad, todas las restricciones son ecuaciones lineales. Entonces tenemos:

1) Si f es una función cóncava, toda solución de las condiciones de primer orden es un maximizador global.

2) Si f es una función convexa, toda solución de las condiciones de primer orden es un minimizador global.

De hecho, sin embargo, hay casos de interés en los que podemos aplicar los criterios de concavidad/convexidad, pero en los que el teorema anterior no es aplicable.

Observad que los supuestos del teorema garantizan que la función h (es decir, el lagrangiano) es cóncava o convexa sobre todo el dominio de las funciones, se cumplan o no las restricciones. Sin embargo, para el problema de optimización que nosotros consideramos tenemos suficiente con el hecho de que la función objetivo sea cóncava o convexa sobre el conjunto factible, siempre que este último sea un conjunto convexo. En este caso, podemos aplicar directamente el teorema local-global, siempre que conozcamos el carácter local de los puntos que solucionan las condiciones de primer orden.

Ejemplo 4.8. En el problema

$$\begin{array}{ll} \max\limits_{(x,y)} & xy \\ \text{s.a.} & x+y=1 \end{array}$$

que hemos visto al principio, es fácil apreciar que las condiciones suficientes de segundo orden para un maximizador local se aplican en el punto $(x^*,y^*,\lambda) = (1/2, 1/2, 1/2)$. Dado que el conjunto factible es convexo, si probamos que la función objetivo es cóncava sobre el conjunto factible, podemos aplicar el teorema local-global para concluir que el punto crítico es un maximizador global. Podemos ver que la función objetivo es cóncava sobre el conjunto factible simplemente expresando, a partir de la restricción, una de las variables en términos de la otra, con lo que nos queda una función cóncava de una variable.

Otra forma de hacerlo sería considerando la forma cuadrática formada a partir de la matriz hessiana de f (evaluada en cualquier punto factible), pero sólo a lo largo de direcciones factibles. En cualquier caso, resulta un ejercicio sencillo, que dejamos al estudiante que esté interesado.

Consideremos ahora la función

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

La matriz hessiana de h es

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

Por lo tanto, h no es globalmente cóncava ni convexa (aunque cuando la restringimos al conjunto factible sea cóncava), por lo que no podemos aplicar el teorema de globalidad ni su corolario.

Cuando nos encontramos con restricciones no lineales y el teorema de globalidad no es aplicable, tenemos un criterio un poco *ad hoc* que puede resultar útil. Este criterio consiste en una aplicación del teorema de Weiertrass; la idea es muy sencilla.

Supongamos que tenemos un problema con restricciones de igualdad que cumple las siguientes condiciones:

- El conjunto factible es compacto (acotado y cerrado).
- Todos los puntos factibles son regulares

Entonces, ya que en los problemas que consideramos aquí la función objetivo siempre es continua (es diferenciable), disponemos de la siguiente información:

El teorema de Weiertrass garantiza la existencia de, al menos, un maximizador global y un minimizador global.

El hecho de que todos los puntos factibles sean regulares garantiza que el método de los multiplicadores de Lagrange dará como puntos críticos cualquier maximizador local y cualquier minimizador local.

La conjunción de los dos últimos resultados nos asegura que necesariamente hay un maximizador global y un minimizador global, que se encuentran entre los puntos críticos. Cuando no hay demasiados puntos críticos, una simple inspección del valor de la función objetivo en cada uno nos permite determinar qué optimizadores son globales y cuáles son sólo locales.

Ejemplo 4.9. Supongamos que queremos encontrar los extremos de $f(x, y) = \frac{2}{3}y^3$ sobre la superficie definida por $g(x,y) = x^2 + y^2 - 1 = 0$. El lagrangiano es $L(x, y, \lambda) = \frac{3}{2}y^3 - \lambda\ (x^2 + y^2 - 1)$, y las condiciones de primer orden:

$$\frac{\partial L}{\partial x} = -2\lambda x = 0$$

$$\frac{\partial L}{\partial x} = 2y^2 - 2\lambda y = 0$$

$$x^2 + y^2 = 1.$$

Hay cuatro soluciones:

$$(x^*, y^*, \lambda^*) = (0,\ 1,\ 1)$$

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

$$(x^*, y^*, \lambda^*) = (1,\ 0,\ 0)$$

$$(x^*, y^*, \lambda^*) = (-1,\ 0,\ 0).$$

Observemos que el teorema de globalidad que hemos dado al principio de este apartado no es aplicable. Para verlo, consideremos el primer punto, $(x^*, y^*, \lambda^*) = (0,1,1)$. Sustituyendo $\lambda = 1$ en el lagrangiano, formamos la función h, que es $h(x, y) = \frac{2}{3}y^3 - x^2 - y^2 + 1$. La matriz hessiana de h es

$$D^2h(x, y) = \begin{bmatrix} -2 & 0 \\ 0 & 4y - 2 \end{bmatrix}$$

y, por lo tanto, la función h no es (globalmente) ni cóncava ni convexa, ya que la matriz es definida negativa en algunos puntos e indefinida en otros. Por lo tanto, no podemos aplicar el teorema. El análisis de los otros puntos da el mismo resultado negativo.

Por otro lado, dado que el conjunto factible es una circunferencia, se trata de un conjunto compacto (acotado y cerrado). El gradiente de la restricción es $\nabla g(x,y) = (2x,2y)$, y sólo puede ser cero si x e y son cero simultáneamente, aspecto que no cumple ningún punto factible; por lo tanto, todos los puntos factibles son regulares. Podemos concluir que, entre los puntos que solucionan las condiciones de primer orden, hay necesariamente un maximizador global y un minimizador global. Evaluando la función objetivo en los puntos críticos encontramos:

$$\begin{aligned} f(0,1) &= 1 \\ f(0,-1) &= -1 \\ f(1,0) &= 0 \\ f(-1,0) &= 0. \end{aligned}$$

Esto nos indica que (0,1) es el maximizador global y (0,–1), el minimizador global. Para hacer un análisis completo del problema, observamos que la función objetivo es una transformación estrictamente creciente de la función $g(x,y) = y$. Por lo tanto, los maximizadores o los minimizadores de f son los mismos que los de q. Esto nos lleva a la conclusión de que, de los cuatro puntos que hemos encontrado, el primero es un maximizador global, y los otros dos no son ni maximizadores ni minimizadores (de hecho, son puntos de silla de la función f).

4.4. Condiciones de segundo orden

Igual que hemos hecho en problemas sin restricción, el análisis de la matriz hessiana de la función objetivo en problemas con restricciones de igualdad nos permite en ocasiones determinar la naturaleza local de los puntos críticos. En este caso también hay otro elemento que debemos tener en cuenta: la restricción.

En problemas sin restricciones, la matriz hessiana evaluada en un punto crítico nos permite saber, cuando ésta es definida positiva o negativa, o es indefinida, si la función crece o decrece en alguna dirección a partir del punto crítico. Cuando hay restricciones no nos interesan todas las direcciones, sino sólo aquéllas en las que hay puntos factibles.

Nota importante

Este apartado incluye material de un nivel más avanzado que el de este curso, por lo que se debe considerar a todos los efectos como material complementario destinado sólo a quien tenga un interés particular en el tema. Aquí presentamos este apartado puramente como referencia para el estudiante interesado. El estudiante de este curso hará bien saltándoselo en una primera lectura del material.

Ejemplo 4.10. Recordemos el problema

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

en el que el único punto crítico es $x^* = y^* = \lambda^* = 1/2$. Si efectuamos la sustitución $y = 1-x$, la función objetivo pasa a ser $h(x) = x(1-x)$, que es una función estrictamente cóncava ($h''(x) = -2<0$, para todo x). Esto quiere decir que las condiciones de segundo orden de este problema se tendrían que cumplir. Sin embargo, la matriz hessiana de la función objetivo es

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

y esta matriz es indefinida, lo cual parece indicar que en algunas direcciones, la función crece, mientras que en otras, la función decrece. Sin embargo, en este problema no debemos considerar cualquier dirección, sino sólo aquellas direcciones en las que hay puntos factibles. Las direcciones factibles vienen dadas por el plano tangente a la restricción, es decir, son aquellas direcciones perpendiculares al vector gradiente $\nabla g(x^*,y^*) = (1,1)$. Con esta restricción es sencillo mostrar que la forma cuadrática generada es definida negativa.

En un problema sin restricciones, la matriz hessiana en un punto crítico nos indica cómo es la curvatura de la función objetivo alrededor de este punto, y esto nos permite conocer el crecimiento o el decrecimiento de la función. Cuando hay restricciones, para saber si la función es creciente o decreciente debemos tener en cuenta no sólo la curvatura de la función objetivo, sino también la curvatura de la restricción. Por este motivo, la matriz hessiana que debemos considerar es la del lagrangiano, que incluye las restricciones además de la función objetivo.

Ejemplo 4.11. Consideremos el problema

$$\begin{array}{ll} \underset{(x,y)}{min} & -x^2+y \\ \text{s.a.} & 2x^2-y=0 \end{array}$$

La restricción es $y = 2x^2$, que, sustituida en la función objetivo, muestra que el problema equivale a minimizar la función x^2, que es una función convexa y, por lo tanto, tiene un mínimo global en $x = 0$. El lagrangiano es $L(x,y,\lambda) = -x^2 + y-\lambda(2x^2 - y)$, y las condiciones de primer orden tienen como única solución: $x^* = 0$, $y^* = 0$ y $\lambda^* = -1$. Sin embargo, la matriz hessiana de la función objetivo es:

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

que es semidefinida negativa, y nos indica que la función decrece o se mantiene constante en cualquier dirección lineal: esto parece indicar que el punto encontrado podría ser un maximizador, pero en ningún caso puede ser un minimizador. Antes hemos visto que el punto es efectivamente un minimizador, y el problema consiste en que tenemos que analizar la curvatura de la función objetivo a lo largo de la dirección cuadrática que define el conjunto factible. La matriz hessiana del lagrangiano respecto a las variables (x,y) es

$$D^2_{(x,y)}L(x,y,\lambda) = \begin{pmatrix} -2-4\lambda & 0 \\ 0 & 0 \end{pmatrix}$$

Teniendo en cuenta que en la solución tenemos $\lambda = -1$, el elemento superior izquierdo de la hessiana es igual a $+2$ en esta solución y, por lo tanto, la posibilidad de un mínimo ya no queda excluida. De hecho, si analizamos la forma cuadrática generada sólo a lo largo de las direcciones factibles, veremos que ésta es definida positiva.

En definitiva, cuando se producen restricciones de igualdad, las condiciones de segundo orden vienen dadas a partir de la matriz hessiana del lagrangiano, evaluando la forma cuadrática correspondiente sólo a lo largo de las direcciones factibles. Observamos que el lagrangiano es función tanto de las variables de decisión como de los multiplicadores, y que para comprobar las condiciones de segundo orden no nos interesa toda la matriz hessiana del lagrangiano, sino sólo aquella parte generada a partir de las derivadas parciales respecto a las variables de decisión.

Denotamos para $D_v^2 L(v, \lambda)$ la **submatriz hessiana** formada por las derivadas parciales segundas del lagrangiano respecto a las variables v, y, dado cualquier vector $h \in \mathrm{R}^n$, sea $D_v^2 L(v, \lambda)$ (h,h) la forma cuadrática generada por aquella submatriz hessiana aplicada al vector h (escribimos el vector h dos veces porque debemos premultiplicar y postmultiplicar la matriz por el vector). Dado un cierto v factible y regular, el plano tangente al conjunto factible que pasa por el punto v viene dado por[1]

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

El punto *v*

Si el punto no es regular, el plano tangente está contenido dentro del conjunto *T(v)*, pero puede no coincidir con éste.

Condiciones suficientes 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). Sea v^* un punto factible que satisface las condiciones de primer orden del método de Lagrange, y sea λ^* el correspondiente multiplicador de Lagrange. Entonces tenemos:

1) Condición suficiente por maximización local. Supongamos que $D_v^2 L(v^*, \lambda^*)$ $(h,h) < 0$ para cada $h \in T(v^*)$ tal que $h \neq 0$. Entonces v^* es un maximizador local estricto.

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

Ejemplo 4.12. Sean $f(x,y) = -x^2 + y$, $g(x,y) = x^2 + y^2 - 1 = 0$. Las condiciones de primer orden nos dan cuatro soluciones:

$$(x, y, \lambda) = \left(0, 1, \frac{1}{2}\right)$$

$$(x, y, \lambda) = \left(0, -1, -\frac{1}{2}\right)$$

$$(x, y, \lambda) = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1\right)$$

$$(x, y, \lambda) = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}, -1\right).$$

El vector gradiente a la restricción es $\nabla g(x,y) = (2x,2y)$, por lo que el plano tangente al conjunto factible en el punto (x,y) está formado por todos aquellos vectores (h_1,h_2), de modo que $x\ h_1 + yh^2 = 0$. Por otro lado, la matriz de derivadas segundas del lagrangiano respecto a (x,y) y la forma cuadrática correspondiente son

$$\begin{pmatrix} -2-2\lambda & 0 \\ 0 & -2\lambda \end{pmatrix} \to Q(h_1, h_2) = -(2+2\lambda)h_1^2 - 2\lambda h_2^2.$$

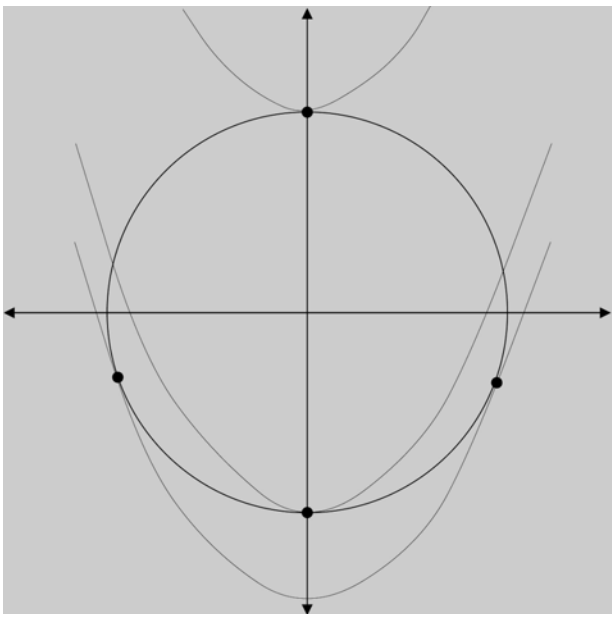
Lo que debemos hacer para cada punto crítico (x, y, λ) es evaluar $Q(h_1,h^2)$ para todos aquellos $(h_1,h^2) \neq (0,0)$, de modo que $xh_1 + yh^2 = 0$.

$$(x,y,\lambda) = \left(0, 1, \tfrac{1}{2}\right), \qquad h_2 = 0 \ \to Q(h_1,h_2) = -3h_1^2 < 0 \ \to Max.$$

$$(x,y,\lambda) = \left(0, -1, -\tfrac{1}{2}\right), \qquad -h_2 = 0 \ \to Q(h_1,h_2) = -h_1^2 < 0 \ \to Max.$$

$$(x,y,\lambda) = \left(\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1\right), \qquad \sqrt{3}h_1 - h_2 = 0 \ \to Q(h_1,h_2) = -6h_1^2 > 0 \ \to Min.$$

$$(x,y,\lambda) = \left(-\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}, -1\right), \qquad -\sqrt{3h_1} - h_2 = 0 \ \to Q(h_1,h_2) = 6h_1^2 > 0 \ \to Min.$$



Parábolas sobre círculo

Las condiciones necesarias correspondientes son similares. Recordad que las condiciones que usamos en la resolución de problemas son las suficientes, no las necesarias, ya que las primeras nos aseguran que un punto crítico determinado es un maximizador o un minimizador local.

Condiciones necesarias de segundo orden

Supongamos que $m \leq n$ y que $f, g_1, g^2 \ldots g_m$ son continuamente diferenciables dos veces (sus derivadas parciales segundas son funciones continuas).

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

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

4.5. Derivación analítica de los resultados

Presentaremos aquí de modo más bien informal una posible derivación de la regla de los multiplicadores de Lagrange y de las condiciones de segundo orden.

Empezamos definiendo el marco en el que trabajaremos. Hay un cierto dominio (abierto) $D \subset \mathbb{R}^n$ y funciones diferenciables dos veces con continuidad $f: D \to \mathbb{R}$ y $g_i : D \to \mathbb{R}$, $1 \leq i \leq m$, donde $m \leq n$. Definimos $X = \{x \in D : g_i(x) = 0, 1 \leq i \leq m\}$. El argumento que damos a continuación se basa en el concepto geométrico del plano tangente a la superficie X en un punto x^*. Para definir esto formalmente, en primer lugar necesitamos otro concepto.

Nota importante

Este apartado incluye material de un nivel más avanzado que el de este curso, por lo que se debe considerar a todos los efectos como material complementario destinado sólo a quien tenga un interés particular en el tema. Aquí presentamos este apartado puramente como referencia para el estudiante interesado. El estudiante de este curso hará bien saltándoselo en una primera lectura del material.

Dado un intervalo real abierto I que contiene el 0, una **trayectoria** sobre X es una función $x: I \to X$ que tiene derivadas segundas continuas.

La noción intuitiva de lo que sería una trayectoria sobre el conjunto X correspondería a la **imagen** de una función que cumpliese las condiciones anteriores.

Dado $x^* \in X$, definimos el **plano tangente** a X en el punto x^* como el conjunto de todas las derivadas $x'(0)$ de todas aquellas trayectorias $x(t)$ sobre X que den como resultado $x(0) = x^*$.

Mediante el uso del teorema de la función implícita, se puede probar que siempre que x^* sea un punto regular, el plano tangente está formado por todos aquellos vectores que son perpendiculares a los vectores gradientes

$$\{\nabla g_1(x^*), \nabla g_2(x^*), ..., \nabla g_m(x^*)\}$$

(que por el supuesto de regularidad son linealmente independientes). Es decir, el plano tangente está formado por todos aquellos vectores v que satisfacen

$$\nabla g_i(x^*) \cdot u = 0, \ 1 \leq i \leq m.$$

En este caso, el subespacio vectorial engendrado por los gradientes de las restricciones forma el plano perpendicular o el plano normal al plano tangente, y los dos subespacios son complementos ortogonales (es decir, engendran todo R^n).

Para fijar las cosas, a partir de ahora supondremos que la función f tiene un máximo local en un punto x^* sobre la superficie X.

Nota

El caso de un mínimo local se analiza de forma similar.

También supondremos que la condición de regularidad se satisface en x^*.

Imaginemos una cierta trayectoria $x(t)$ sobre X que pasa por el punto x^*. Entonces la función de una variable $h(t) = f[x(t)]$ tiene un maximizador local en el punto interior $t = 0$ y, por lo tanto, sabemos que $h'(0) = 0$. Aplicando la regla de la cadena:

$$h'(0) = \nabla f[x(0)] \cdot x'(0) = 0.$$

Esta condición de primer orden implica que el vector gradiente $\nabla f[x(0)]$ es perpendicular al plano tangente, y (dada la condición de regularidad) esto quiere decir que este vector gradiente pertenece al plano normal y, por lo tanto, es combinación lineal de los vectores gradientes de las restricciones, es decir:

$$\nabla f[x(0)] = \lambda_1 \nabla g_1[x(0)] + \lambda^2 \nabla g^2[x(0)] + ... + \lambda_m \nabla g_m[x(0)].$$

Esto es lo que nosotros hemos llamado antes **condición de tangencia** de la regla de los multiplicadores de Lagrange.

Para estudiar la condición suficiente de segundo orden, lo que tenemos que hacer es analizar la derivada segunda $h''(0)$ de la función de una variable h. Aplicando de nuevo la regla de la cadena, obtenemos:

$$h''(0) = D^2 f[x(0)](x'(0), x'(0)) + \nabla f[x(0)] \cdot x''(0),$$

donde, como antes, denotamos para $D^2 f[x(0)](x'(0), x'(0))$ la forma cuadrática generada a partir de la matriz hessiana de f y el vector $x'(0)$.

Es interesante detenernos aquí e interpretar los términos que aparecen en la expresión de la función h'', que describe la curvatura de h. El primer término es la matriz hessiana de f evaluada a lo largo de la dirección tangente dada por x'; por lo tanto, retoma la influencia de la curvatura de f. En el segundo término nos encontramos con la derivada segunda de la trayectoria x'', que captura la influencia de la curvatura de la superficie factible sobre h.

Por otro lado, dado que para todo $t \in I$ se cumple que:

$$\sum_{i=1}^{m} \lambda_i \, g_i[x(t)] = 0,$$

diferenciamos dos veces esta función constante y, evaluándola en el punto 0, obtenemos:

$$\sum_{i=1}^{m} \lambda_i \nabla g_i[x(0)] \cdot x''(0) + \sum \lambda_i D^2 g_i[x(0)](x'(0), x'(0)) = 0$$

Observamos que la condición de tangencia implica que

$$\nabla f[x(0)] \cdot x''(0) = -\sum_{i=1}^{m} \lambda_i \nabla g_i[x(0)] \cdot x''(0)$$

y, por tanto,

$$\nabla f[x(0)] \cdot x''(0) = \sum_{i=1}^{m} \lambda_i D^2 g[x(0)](x'(0), x'(0))$$

Esta última igualdad es interesante, ya que nos enseña cómo podemos expresar la curvatura de la superficie factible, retomada por la derivada segunda de la trayectoria x'', en términos de las matrices hessianas de las restricciones aplicadas a la dirección tangente x'. Sustituyendo todos estos factores en la expresión de la derivada segunda de h que hemos encontrado antes, obtenemos finalmente

$$h''(0) = D^2 f[x(0)](x'(0), x'(0)) - \sum_{i=1}^{m} \lambda_i D^2 g_i[x(0)](x'(0), x'(0)).$$

De este modo, vemos que la curvatura de h viene dada por la matriz hessiana del lagrangiano, aplicada a la dirección tangente x'. Recordemos que el lagrangiano de este problema es $L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x)$, por lo que, si designamos $D_x^2 L(x,\lambda)$ a la submatriz hessiana del lagrangiano formada por las derivadas parciales de segundo orden respecto a las variables x, nos queda

$$h''(0) = D_x^2 L[x(0), \lambda](x', (0), x'(0)).$$

A partir de esta expresión podemos derivar tanto la condición necesaria como la condición suficiente de segundo orden. Por ejemplo, la condición suficiente exigirá que la forma cuadrática definida a partir del lagrangiano sea negativa en $x'(0)$ para cualquier trayectoria que satisfaga $x(0) = x^*$ y $x'(0) \neq 0$. Dicho de otro modo, que la forma cuadrática definida por $D_x^2 L(x^*,\lambda)$ tome valores negativos para todos los vectores (diferentes del 0) que pertenezcan al plano tangente.

5. Optimización con restricciones de desigualdad

5.1. Una primera aproximación a problemas con desigualdades

Consideremos el siguiente problema de minimización:

$$\begin{array}{ll} \underset{(x,y)}{min} & (x-1)^2+(y-1)^2 \\ \text{s.a.} & x^2+y^2 \leq 8. \end{array}$$

Supongamos que un cierto punto, llamémosle (a,b), es un minimizador local del problema (por cierto, el teorema de Weiertrass nos asegura que hay por lo menos un minimizador global). Entonces pueden suceder dos cosas: o bien $a^2 + b^2 = 8$ o bien $a^2 + b^2 < 8$. En el primer caso, diremos que la restricción es efectiva en el punto (a,b). Si la restricción es efectiva en (a,b), significa que (a,b) también es solución del problema con una restricción de igualdad

$$\begin{array}{ll} \underset{(x,y)}{min} & (x-1)^2+(y-1)^2 \\ \text{s.a.} & x^2+y^2 = 8. \end{array}$$

y, por lo tanto, sabemos que hay un cierto λ^* que, si $f(x,y) = (x-1)^2 + (y-1)^2$ y $g(x,y) = x^2 + y^2 - 8$, entonces el vector (a,b, λ^*) soluciona el sistema de ecuaciones

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \lambda\ \frac{\partial g}{\partial x}(x,y) \\ \frac{\partial f}{\partial y}(x,y) &= \lambda\ \frac{\partial g}{\partial y}(x,y) \\ g(x,y) &= 0. \end{aligned} \tag{5.1}$$

Supongamos ahora que la restricción no es efectiva en (a,b), es decir, que $a^2 + b^2 < 8$. Entonces (a,b) es solución del problema de optimización

$$\begin{array}{ll} \underset{(x,y)}{min} & (x-1)^2+(y-1)^2 \\ \text{s.a.} & x^2+y^2 < 8. \end{array}$$

Este problema no parece demasiado distinto del que teníamos originalmente, pero hay una diferencia muy importante: al ser la restricción una desigualdad estricta, el conjunto de puntos factibles es abierto, es decir, todo punto factible está rodeado en cualquier dirección por otros puntos factibles.

Si repasáis el razonamiento que hicimos en problemas sin restricciones para justificar que en aquel caso todas las derivadas parciales deben ser cero, veréis que también se aplica al caso que estamos considerando, ya que lo que resulta esencial en este razonamiento es que, en cualquier dirección a partir del punto crítico, haya otros puntos factibles. En definitiva, si (a,b) soluciona el último problema que hemos escrito, entonces debe ser solución del siguiente sistema de relaciones:

$$\begin{aligned} &\frac{\partial f}{\partial x}(x, y) = 0 \\ &\frac{\partial f}{\partial y}(x, y) = 0 \\ &g(x, y) < 0. \end{aligned} \tag{5.2.}$$

Por lo tanto, cualquier minimizador local del problema que hemos planteado originalmente satisface las relaciones (5.1) o las relaciones (5.3). Una forma compacta de expresar las condiciones que deben satisfacer todos estos candidatos es la siguiente: buscaremos aquellos vectores (x,y,λ) que satisfagan

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \lambda \ \frac{\partial g}{\partial x}(x, y) \\ \frac{\partial f}{\partial y}(x, y) &= \lambda \ \frac{\partial g}{\partial y}(x, y) \\ g(x, y) &= 0 \end{aligned}$$

$$g(x,y) < 0 \text{ implica que } \lambda = 0.$$

Observad que la última condición hace que, cuando un punto satisface la desigualdad de forma estricta, la λ que le corresponde sea cero y que, por lo tanto, las condiciones de tangencia consistan sencillamente en igualar las derivadas parciales de f a cero.

Las ecuaciones que nos quedan son

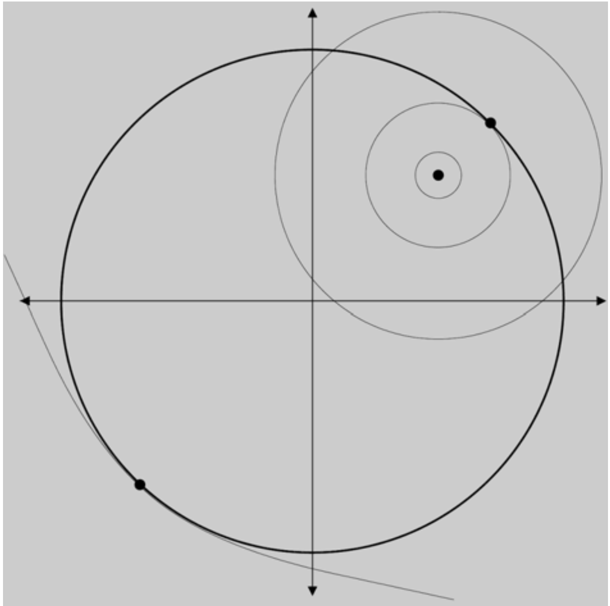
$$\begin{aligned} 2x-2 &= 2\lambda x \\ 2y-2 &= 2\lambda y \\ x^2 + y^2 &\le 8 \end{aligned}$$

$$x^2 + y^2 < 8 \text{ implica que } \lambda = 0.$$

Las soluciones en términos de (x,y,λ) son

$$(2, 2, 1/2), \ (-2, -2, 3/2) \ \text{ y } \ (1, 1, 0).$$

En definitiva, tenemos tres candidatos en los que las variables de decisión valen (1, 1), (2, 2) y (–2, –2). ¿Podemos decir algo más sobre el problema? Debéis notar que el conjunto factible es convexo, ya que se trata de un círculo (sólido), y la función objetivo f es convexa; dado que (1, 1) es un punto crítico de f (todas sus derivadas parciales son 0), tenemos un teorema que nos asegura que (1, 1) es un minimizador global de f sobre su dominio. Es decir, la solución del problema formulado originariamente es el punto (1, 1). Para ver qué son los otros puntos, nos ayudaremos con un gráfico.



Restricciones de desigual

Puntos críticos de...

...
$f(x,y) = (x-1)^2 + (y-1)^2$ sobre la superficie $x^2 + y^2 \leq 8$.

Por ejemplo, consideremos el punto (2, 2). Si desde este punto nos movemos hacia el origen, f toma valores inferiores, mientras que si nos movemos a lo largo de la circunferencia, f toma valores superiores. O sea, que (2, 2) no es ni maximizador ni minimizador local. ¿Por qué nos ha salido como punto crítico? El motivo es que para encontrarlo hemos aplicado el método de Lagrange tomando como restricción sólo la circunferencia, sobre la que el punto es, efectivamente, un minimizador local. Por otro lado, a partir del gráfico podemos ver que el punto (–2,–2) es un maximizador global.

En conclusión, combinando los métodos de optimización sin restricciones y de optimización con restricciones de igualdad, podemos solucionar los problemas con desigualdades, pero encontramos la dificultad de que nos aparecen puntos críticos no válidos. Querríamos afinar un poco más nuestro análisis y ser capaces de descartar estos puntos adicionales. El resultado de este afinamiento es lo que denominamos **método de Kuhn y Tucker**.

5.2. El método de Kuhn-Tucker

En el ejemplo anterior, los puntos críticos no válidos han aparecido cuando hemos impuesto que la restricción se cumpliese con una igualdad. El método de Kuhn y Tucker se deriva de un análisis más detallado de los puntos críticos que aparecen cuando imponemos que algunas de las restricciones se satisfagan con igualdad.

El fundamento de este método consiste en, dado un cierto candidato, utilizar los gradientes de las restricciones efectivas en este punto para saber hacia dónde se encuentra el conjunto factible, así como en utilizar el gradiente de la función objetivo para saber en qué dirección hay puntos que nos acercan más a nuestro objetivo (de maximizar o minimizar); si se da coincidencia entre puntos factibles y puntos en los que nos acercamos más a nuestro objetivo, entonces el candidato no puede ser un extremo local.

Consideremos de nuevo el ejemplo del apartado 5.1. Recordemos que hemos definido $f(x,y) = (x-1)^2 + (y-1)^2$ y $g(x,y) = x^2 + y^2 - 8$. Consideremos el punto crítico $(x,y) = (2,2)$, en el que el λ correspondiente es 1/2, ya que $\nabla f(2,2) = (2,2)$ y $\nabla g(2,2) = (4,4)$ (en los puntos críticos con una restricción de igualdad, se satisface $\nabla f = \lambda \nabla g$). El hecho de que λ sea positivo indica que los dos gradientes, además de ser colineales, apuntan en el mismo sentido. Como la restricción del problema original es $g(x,y) \leq 0$, sabemos que, alrededor del punto (2, 2) (en el cual g toma exactamentet el valor 0), encontraremos puntos factibles si nos movemos en la dirección opuesta a la que marca el gradiente.

Por otro lado, el punto (2, 2) será un mínimo local de la función f siempre que no haya puntos factibles a su alrededor en los que la función tome un valor inferior. Lo que vemos es, de este modo, que (2, 2) no puede ser un minimizador local, ya que, justo en la dirección en la que sabemos que la función toma un valor inferior (la dirección opuesta a la que señala el gradiente), hay puntos factibles. Un análisis similar muestra que el punto (–2, –2) no puede ser un minimizador local del problema.

La clave del argumento anterior consiste en ir comprobando, para cada candidato, si hay puntos factibles en la dirección de disminución del valor de la función objetivo; si es éste el caso, podemos descartar este candidato, ya que no puede tratarse de un minimizador local. Además, la comprobación que debemos hacer se limita a observar hacia dónde apuntan los gradientes (evaluados en el punto en cuestión), teniendo en cuenta la forma como hemos escrito la desigualdad. Decimos que un problema de optimización con dos variables y una restricción de desigualdad está escrito de forma **normalizada**, si definimos las funciones f y g de tal modo que el problema llegue a ser, según si se trata de maximizar o de minimizar:

$$\begin{array}{ll} \max\limits_{(x,y)} & f(x,y) \\ \text{s.a.} & g(x,y) \leq o \end{array} \qquad\qquad \begin{array}{ll} \min\limits_{(x,y)} & f(x,y) \\ \text{s.a.} & g(x,y) \geq 0 \end{array}$$

Es decir, en el caso de maximización tenemos que escribir la desigualdad como $\leq$ y en el de minimización, como $\geq$.

Las condiciones de Kuhn-Tucker también serán diferentes según queramos maximizar o minimizar.

Maximización	Minimización
$\nabla f(x,y) = \lambda \nabla g(x,y)$, $g(x,y) \leq 0$, $\lambda \geq 0$, $g(x,y)<0$ implica $\lambda = 0$.	$\nabla f(x,y) = \lambda \nabla g(x,y)$ $g(x,y) \geq 0$ $\lambda \geq 0$, $g(x,y)>0$ implica $\lambda = 0$.

En la práctica,...

... lo más sencillo es escribir los problemas de maximización y de minimización de tal modo que la comprobación se reduzca a verificar que los gradientes de la función objetivo y de la restricción apunten en la misma dirección, es decir, que el λ correspondiente sea positiva.

El requerimiento de que $\lambda = 0$ cuando la restricción no es efectiva recibe el nombre de **condición de exclusión o de complementariedad**.

Una forma útil de escribir las restricciones de factibilidad, de no negatividad de la λ y de exclusión es, en el caso de la maximización:

$$\lambda \geq 0,\ g(x,y) \leq 0,\ \lambda g(x,y) = 0.$$

Esto es más sencillo que la implicación que escribíamos antes, aunque obviamente es lo mismo (teniendo en cuenta las otras condiciones).

Cuando hay muchas variables y muchas restricciones, todo lo que acabamos de razonar se generaliza. Sin embargo, debemos tener en cuenta que, cuando hay restricciones de desigualdad, un detalle cambia con respecto a cuando todas las restricciones son de igualdad. De hecho, es fácil imaginar problemas bien definidos con dos variables y un número arbitrario de restricciones (no triviales) de desigualdad: por ejemplo, desigualdades lineales que dan lugar como conjunto factible a un polígono con un número de caras dado (tan grande como se quiera).

Empezamos la parte más formal de este apartado con definiciones matemáticas de los conceptos que hemos ido viendo en los ejemplos anteriores. A partir de aquí, supongamos que tenemos dado un cierto problema de optimización con restricciones de desigualdad (incluyendo también el caso en el que hay tantas igualdades como desigualdades).

Decimos que una restricción es **efectiva** en un punto si éste la satisface con igualdad.

Ejemplo 5.1. La restricción $x^2 + y^2 \leq 8$ es efectiva en el punto (2, 2), pero no lo es en el punto (1, 1).

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 el punto)

Ejemplo 5.2. En el conjunto factible definido por:

$$(x-1)^2 + (y-2)^2 \leq 1$$

$$(x-1)^2 - y \leq -1,$$

todos los puntos son regulares excepto (1, 1).

Podéis comprobar...

... que todos los puntos son regulares excepto (1, 1). Es fácil verificar mediante una representación gráfica que las dos restricciones son tangentes en el punto (1, 1).

Decimos que un problema de maximización con n variables y m restricciones de desigualdad está escrito de forma **normalizada** cuando lo podemos expresar como:

$$\begin{aligned} &\max_{v \in IR^n} f(v) \\ \text{s.a.}\quad &g_1(v) \leq 0 \\ &g_2(v) \leq 0 \\ &\ldots\ldots\ldots \\ &g_m(v) \leq 0 \end{aligned} \qquad (5.3.)$$

Decimos que un problema de minimización con n variables y m restricciones de desigualdad está escrito de forma **normalizada** cuando lo podemos expresar como:

$$\begin{aligned} &\min_{v \in IR^n} f(v) \\ \text{s.a.}\quad &g_1(v) \geq 0 \\ &g_2(v) \geq 0 \\ &\ldots\ldots\ldots \\ &g_m(v) \geq 0 \end{aligned} \qquad (5.4.)$$

Consideremos el problema normalizado de maximización 5.3. Supongamos que las funciones f, g_1,g^2...g_m son diferenciables. Las **condiciones de Kuhn-Tucker** del problema son

$$\nabla f(v) = \lambda_1 \nabla g_1(v) + \lambda_2 \nabla g^2(v) + \ldots + \lambda_m \nabla g_m(v) \rightarrow \text{Tangencia}$$

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

Consideremos el problema normalizado de minimización 5.4. Supongamos que las funciones f, g_1,g^2...g_m son diferenciables. Las c**ondiciones de Kunh-Tucker** del problema son

$$\nabla f(v) = \lambda_1 \nabla g_1(v) + \lambda_2 \nabla g^2(v) + \ldots + \lambda_m \nabla g_m(v) \rightarrow \text{Tangencia}$$

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

Tal y como sucede en el caso de restricciones de igualdad, una forma conveniente de expresar las condiciones de Kuhn y Tucker es hacerlo mediante el lagrangiano del problema de optimización.

El **lagrangiano**, o función de Lagrange, del problema normalizado de maximización 5.3, o del problema normalizado de minimización 5.4, es la función de $n + m$ argumentos:

$$L(v_1,v^2...v^n, \lambda_1, \lambda_2... \lambda_m) =$$

$$f(v_1,v^2...v^n)- \lambda_1 g_1(v_1,v_2...v_n) - ...$$

$$... - \lambda_2\, g_2(v_1,v_2...v_n) - ... - \lambda_m\, g_m\, (v_1,v_2...v_n).$$

Tanto si un problema es de maximización como si es de minimización, las condiciones de tangencia son las que resultan de igualar a cero las derivadas parciales del lagrangiano respecto a las variables de decisión.

Ejemplo 5.3. Consideremos el problema de minimización

$$\begin{array}{ll} \min\limits_{(x,y)} & xy \\ \text{s.a.} & x^2 + y^2 \leq 2. \end{array}$$

Para encontrar las condiciones de Kuhn y Tucker, lo primero que hacemos es escribirlo de forma normalizada:

$$\begin{array}{ll} \min\limits_{(x,y)} & xy \\ \text{s.a.} & -x^2 - y^2 + 2 \geq 0 \end{array}$$

Por lo tanto: $f(x,y) = xy$, $g(x,y) = -x_2 - y_2 + 2$. A continuación definimos el lagrangiano:

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

Las condiciones de Kuhn y Tucker son:

$$\begin{array}{rll} \dfrac{\partial L}{\partial x} = y + 2\lambda x = 0 & & \text{(a)} \\ \dfrac{\partial L}{\partial y} = x + 2\lambda y = 0 & & \text{(b)} \\ \lambda \geq 0 & & \text{(c)} \\ x^2 + y^2 \leq 2 & & \text{(d)} \\ \lambda(- x^2 - y^2 + 2) = 0 & & \text{(e)} \end{array}$$

Para resolver estas condiciones, es típico empezar por las condiciones de tangencia. Igualando (*a*) y (*b*) obtenemos:

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

Ahora tenemos dos posibilidades: o bien $x = y$, o bien $\lambda = 1/2$. Si exploramos todas estas posibilidades, llegamos a una solución o bien a una contradicción. Empecemos por la primera:

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

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

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

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

Las condiciones de exclusión dicen que o bien una restricción es efectiva, o bien el multiplicador es cero. En general, sólo una de las dos cosas es cierta, aunque las dos cosas pueden ser ciertas simultáneamente. De hecho, estos casos en ocasiones pueden dar problemas, como ocurre cuando queremos hacer un análisis de sensibilidad.

Llamaremos **soluciones degeneradas** aquéllas en las que se da esta circunstancia.

Ejemplo 5.4. Consideremos el programa de maximización:

$$\begin{aligned} &\max_{(x,y)} \; xy \\ &\text{s.a.} \quad x + y \leq 2 \\ &\qquad\quad\; y \geq 1. \end{aligned}$$

Normalizamos:

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

El lagrangiano es $L(x,y,\lambda_1,\lambda_2) = xy - \lambda_1(x + y - 2) - \lambda_2(-y + 1)$, y las condiciones de Kuhn-Tucker:

$$\begin{array}{llll} \dfrac{\partial L}{\partial x} = y - \lambda_1 & = 0 & & \text{(a)} \\ \dfrac{\partial L}{\partial y} = x - \lambda_1 + \lambda_2 & = 0 & & \text{(b)} \\ \lambda_1 \geq 0, & x + y \leq 2, & \lambda_1(x + y - 2) = 0 & \text{(c)} \\ \lambda_2 \geq 0, & y \geq 1, & \lambda_2(y - 1) = 0 & \text{(d)} \end{array}$$

Para resolver estas condiciones, empezamos con:

$$(a) \rightarrow \lambda_1 \geq 1 > 0 \xrightarrow{(c)} x + y = 2$$

Por otro lado, (a) y (b) implican que $-x + y = \lambda_2$. Sumando esta igualdad y la que acabamos de obtener, tenemos $y = 1 + \frac{\lambda_2}{2}$. Observamos ahora que esta igualdad y la condición de exclusión de (d) sólo pueden ser válidas si $y = 1$ y $\lambda_2 = 0$. Finalmente, a partir de esto obtenemos que $x = 1$ y $\lambda_1 = 1$.

Observemos que...

... la restricción $y \geq 1$ es efectiva en la solución, pero el multiplicador correspondiente es cero. Se trata de una solución degenerada, aunque una representación gráfica ayuda a ver que, de hecho, el punto encontrado es un maximizador global.

5.3. Justificación alternativa de la no negatividad de los multiplicadores

Supongamos que estamos considerando un problema de maximización escrito de la forma

$$\begin{array}{ll} \max\limits_{v} & f(v) \\ \text{s.a.} & g(v) \leq 0. \end{array}$$

Y consideremos un punto $\bar{v}$ que satisface las condiciones de Lagrange cuando imponemos igualdad en la restricción, es decir $\nabla f(\bar{v}) = \lambda \nabla_g(\bar{v})$.

Lo que ahora queremos hacer es considerar qué direcciones son factibles alrededor de $\bar{v}$. En los problemas con restricciones de igualdad ya hemos visto que las direcciones a lo largo del plano tangente a la restricción son factibles; es decir, aquellas direcciones h que satisfacen

$$\nabla_g(\bar{v}) \cdot h = 0.$$

Sin embargo, la restricción es $g(v) \leq 0$ y, por lo tanto, también serán factibles aquellas direcciones que apunten hacia donde la función g disminuye de valor (sabemos que $g(\bar{v}) = 0$), además de las direcciones en las que g se mantiene constante (que son las que constan sobre el plano tangente). La aproximación lineal (es decir, la aproximación de Taylor de primer orden) a la función g alrededor de $\bar{v}$ nos muestra que las direcciones factibles vendrán dadas por todos aquellos vectores h que satisfacen

$$\nabla g(\bar{v}) \cdot h \leq 0.$$

A pesar de todo, si existe alguna dirección factible h tal que

$$\nabla f(\bar{v}) \cdot h \leq 0,$$

entonces la condición necesaria de maximización estará garantizada siempre que exijamos que $\lambda \geq 0$.

5.4. Condiciones de primer orden

Tal y como sucede en el caso de restricciones de igualdad, y exactamente por los mismos motivos, las condiciones de Kuhn y Tucker serán necesarias siempre que los maximizadores o los minimizadores locales sean puntos regulares.

Condiciones de primer orden

Supongamos que v^* es un maximizador local del problema 5.3 o un minimizador local del problema 5.4, y supongamos adicionalmente que las funciones $f, g_1, g^2 \dots g_m$ son diferenciables con continuidad. Entonces, si v^* es un punto regular, este punto satisface las condiciones de Kuhn y Tucker del problema de optimización respectivo.

Es decir, cuando todos los puntos factibles son regulares, podemos eliminar como candidatos a maximizadores o minimizadores todos aquellos puntos que no satisfacen las condiciones de Kuhn y Tucker.

Ejemplo 5.5. En el problema

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

hemos encontrado que las soluciones de las condiciones de Kuhn y Tucker son los puntos en los que (x, y, λ) valen $(0, 0, 0)$, $(1, -1, 1/2)$ y $(-1, 1, 1/2)$. En el apartado sobre restricciones en forma de igualdad vimos que todos los puntos que constan sobre la circunferencia son regulares. Por otro lado, cualquier punto en el que la restricción no es efectiva es regular, ya que la regularidad es una condición que se debe satisfacer sólo en las restricciones efectivas. En conclusión, todos los puntos factibles son regulares. El teorema anterior implica que, si hay un minimizador local, éste es uno de los tres puntos que acabamos de escribir.

El teorema que hemos visto no dice que, si las condiciones de Kuhn y Tucker tienen soluciones, todas sean la solución del problema de optimización, ni siquiera que este problema tenga alguna. Sólo afirma que las condiciones de Kuhn y Tucker detectarán cualquier maximizador local o cualquier minimizador del mismo tipo.

Ejemplo 5.6. Consideremos el problema (normalizado) de minimización

$$\begin{aligned} &\min_{(x,y)} \; y \\ &\text{s.a.} \quad -x^3 + y \geq 0 \end{aligned}$$

El lagrangiano es $L(x,y,\lambda) = y - \lambda(-x^3 + y)$, y las condiciones de Kuhn y Tucker

$$\frac{\partial L}{\partial x} = 3\lambda x^2 = 0$$

$$\frac{\partial L}{\partial y} = 1 - \lambda = 0$$

$$\lambda \geq 0, \qquad y \geq x^3, \qquad \lambda(-x^3 + y) = 0$$

Es fácil darse cuenta de que la única solución es $(x, y, \lambda) = (0, 0, 1)$, y que todos los puntos factibles son regulares. ¿Quiere esto decir que hemos encontrado la solución del problema de minimización? Observad que, tomando valores negativos muy grandes de la variable x, podemos hacer que la variable y (la función objetiva) sea tan pequeña como queramos y que, por lo tanto, satisfaga la restricción. Esto significa que la función objetivo no es acotada inferiormente y, de este modo, que no tiene un valor mínimo.

5.5. Condiciones de segundo orden

Las condiciones locales de segundo orden no son más que la generalización natural de lo que sucede en el caso de restricciones de igualdad, con una pequeña calificación para el caso de suficiencia. En el caso de restricciones de desigualdad, las restricciones que no sean efectivas no cuentan y las podemos ignorar en lo que respecta a condiciones locales de segundo orden. En lo que respecta a las restricciones efectivas, debemos considerar las direcciones factibles a las que dan lugar. Hay dos tipos de direcciones factibles: las que están sobre el plano tangente y las que van hacia el interior del conjunto factible. De las direcciones factibles que van hacia el interior del conjunto factible ya nos ocupamos al exigir que los multiplicadores sean no negativos. Por lo tanto, la condición local de segundo orden es análoga al caso con restricciones de igualdad, tomando sólo las restricciones efectivas.

Nota importante

Este apartado incluye material de un nivel más avanzado que el de este curso, por lo que se debe considerar a todos los efectos como material complementario destinado sólo a quien tenga un interés particular en el tema. Aquí presentamos este apartado puramente como referencia para el estudiante interesado. El estudiante de este curso hará bien saltándoselo en una primera lectura del material.

Condiciones necesarias de segundo orden

Supongamos que las funciones $f, g_1, g_2 \ldots g_m$ son continuamente diferenciables dos veces. Entonces:

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

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

Condiciones suficientes de segundo orden

Supongamos que las funciones $f, g_1, g^2 \ldots g_m$ son continuamente diferenciables dos veces, y que existe λ^*, de modo que (v^*, λ^*) satisfacen las condiciones de Kuhn y Tucker. Entonces:

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

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

Observad que...

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

5.6. Criterios de globalidad

Una cuestión más restrictiva es saber bajo qué circunstancias todo punto que cumpla las condiciones de Kuhn y Tucker es necesariamente una solución del problema de optimización. Esencialmente, requeriremos que el conjunto factible sea convexo y que la función objetivo sea cóncava o convexa (o generalizaciones), según queramos maximizar o minimizar. Un resultado mucho más restrictivo de lo necesario, pero extremadamente sencillo de demostrar y de comprobar en la práctica, es el siguiente.

Teorema

1) Consideremos el problema normalizado de maximización 5.3. Supongamos que las funciones $f, g_1, g_2, \ldots g_m$ son funciones convexas. Entonces, todo punto que solucione las condiciones de Kuhn y Tucker es un maximizador global del problema.

2) Consideremos el problema normalizado de minimización 5.4. Supongamos que las funciones $f, g_1, g_2 \ldots g_m$ son diferenciables. Adicionalmente, supongamos que f es una función convexa y que $g_1, g_2 \ldots g_m$ son funciones cóncavas. Entonces, todo punto que solucione las condiciones de Kuhn y Tucker es un minimizador global del problema.

Ejemplo 5.7. Consideremos el problema

$$\max_{(x,y)} \; x - y + xy - \frac{1}{4}x^4 - \frac{1}{2}x^2 - y^2$$

$$\begin{aligned} \text{s.a.} \quad x^2 + y^4 &\leq 4 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Normalicemos

$$\underset{(x,y)}{max}\ x - y + xy - \frac{1}{4}x^4 - \frac{1}{2}x^2 - y^2$$

$$\begin{aligned} \text{s.a.}\ \ & x^2 + y^4 - 4 \leq 0 \\ & -x \leq 0 \\ & -y \leq 0 \end{aligned}$$

Definamos

$$f(x, y) = x - y + xy - \frac{1}{4}x^4 - \frac{1}{2}x^2 - y^2$$

$$g_1(x,y) = x^2 + y^4 - 4$$

$$g_2(x,y) = -x$$

$$g_3(x,y) = -y$$

Es una cuestión rutinaria comprobar, utilizando las respectivas matrices hessianas, que f es una función cóncava y g_1,g_2 y g_3 son funciones convexas. Por lo tanto, el teorema anterior implica que cualquier solución de las condiciones de Kuhn y Tucker es automáticamente un maximizador global.

Ejemplo 5.8. Consideremos el problema

$$\underset{(x,y,z)}{min}\ x + y + z + e^x + y^2 + z^4$$

$$\begin{aligned} \text{s.a.}\ \ & x + y + z \geq -2 \\ & x^2 + y^2 + z^2 \leq 8. \end{aligned}$$

Normalicemos

$$\underset{(x,y,z)}{min}\ x + y + z + e^x + y^2 + z^4$$

$$\begin{aligned} \text{s.a.}\ \ & x + y + z + 2 \geq 0 \\ & -x^2 - y^2 - z^2 + 8 \geq 0. \end{aligned}$$

Definamos

$$f(x,y,z) = x + y + z + e^x + y^2 + z^4$$

$$g_1(x,y,z) = x + y + z + 2$$

$$g_2(x,y,z) = -x^2 - y^2 - z^2 + 8.$$

También es algo rutinario comprobar, utilizando las respectivas matrices hessianas, que f es una función convexa y g_1, g_2 son funciones cóncavas. Por lo tanto, el teorema anterior implica que cualquier solución de las condiciones de Kuhn y Tucker es automáticamente un minimizador global.

El resultado del teorema anterior, que nos servirá para comprobar en muchos casos la validez de los puntos obtenidos mediante las condiciones de Kuhn y

Tucker, se deriva de forma bastante inmediata de los resultados que caracterizan las funciones cóncavas y las funciones convexas. A continuación presentamos una demostración de ello.

Demostración del teorema

Probaremos sólo el caso de maximización. Supongamos que (v^*, 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 $L(v, \lambda^*) = f(v) - \sum_{i=1}^{m} \lambda_i^* \, g_i(v)$ es una función cóncava en la variable v, ya que $\lambda^* \geq 0$. Por lo tanto, dado cualquier punto factible v tenemos:

$$L(\bar{v}, \lambda^*) - L(v^*,\lambda^*) \leq \nabla_v L(v^*,\lambda^*)\ (\bar{v} - v^*).$$

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

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

Substituyendo en las definiciones de los lagrangianos respectivos, tenemos:

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

Sin embargo, la condición de exclusión implica que, para todo i, $\lambda_i^* g_i(v^*) = 0$ y, por lo tanto,

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

También sabemos que, para todo $\lambda_i^* \geq 0$ y $g_i(\bar{v}) \leq 0$, o sea que $\lambda^* g_i(\bar{v}) \leq 0$ (recordemos que v no tiene por qué satisfacer la condición de KT y que, por lo tanto, el último producto no tiene por qué ser cero), así que:

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

De este modo, la función objetivo toma un valor mayor en v^* que en $\bar{v}$. Como v era un punto factible cualquiera, podemos concluir que v^* es un maximizador global.

El hecho de que las funciones que definen las restricciones sean cóncavas o convexas, según minimicemos o maximicemos, garantiza que el conjunto factible siempre es un conjunto convexo.

Para generalizar este resultado, necesitaremos generalizar los conceptos de concavidad y de convexidad.

Sea X un subconjunto convexo de R^n.

1) Decimos que la función $f : x \to \mathrm{R}$ **es cuasicóncava** si, para todo $v \in X$, el contorno superior $\{w \in X : f(w) \geq f(v)\}$ es un conjunto convexo.

2) Decimos que la función $f : x \to \mathrm{R}$ **es cuasiconvexa** si, para todo $v \in X$, el contorno inferior $\{w \in X : f(w) \leq f(v)\}$ es un conjunto convexo.

Cuando f es una función continua, que es el único caso que nosotros consideramos, la definición es equivalente si la hacemos en términos de los contornos definidos con desigualdades débiles o estrictas.

En general, comprobar si una función es cuasicóncava o cuasiconvexa puede llegar a ser una tarea algebraicamente bastante compleja. La forma más sencilla de hacerlo es mediante la siguiente proposición, que también nos sirve para entender el origen de la terminología.

Podéis...

... comprobar la validez de esta proposición a partir de las definiciones algebraicas de funciones cóncavas y de funciones convexas.

Proposición 4 *Toda función cóncava es cuasicóncava y toda función convexa es cuasiconvexa*.

Las equivalencias siguientes son básicamente reescrituras de las definiciones respectivas.

Proposición 5. Sea X un subconjunto convexo de R^n.

1) *La función $f : X \to \mathrm{R}$ es cuasicóncava si, y sólo si, para todo v y w de X, y para todo* $0 \leq t \leq 1$, *se cumple que* $f[tv + (1 - t)w] \geq min\{f(v), f(w)\}$.

2) *La función $f : X \to \mathrm{R}$ es cuasiconvexa si, y sólo si, para todo v y w de X, y para todo* $0 \leq t \leq 1$, *se cumple que* $f[tv + (1-t)w] \leq max\{f(v), f(w)\}$.

Recordemos que dos problemas de optimización son equivalentes siempre que tengan el mismo conjunto factible y que las respectivas funciones objetivo sean transformaciones estrictamente crecientes una de otra. El resultado siguiente nos indica que, si queremos estudiar problemas que satisfagan la condición de globalidad que hemos expuesto al principio de este apartado, no podemos evitar considerar también problemas en los cuales la función objetivo sea cuasicóncava o cuasiconvexa.

Proposición 6. *Toda transformación estrictamente creciente de una función cuasicóncava resulta en una nueva función cuasiconvexa; en particular, una transformación estrictamente creciente de una función convexa es una función cuasiconvexa* (*no necesariamente convexa*).

Ejemplo 5.9. La función $f(x) = x^2$, definida sobre $X = [0,\infty)$, es cuasicóncava, pero no es cóncava. La comprobación de la cuasiconcavidad es sencilla a partir de la misma definición, ya que para todo número real $t \geq 0$, el contorno superior de nivel t es el intervalo $[\sqrt{t}, +\infty)$. Dado que para todo $x \in X$, $f''(x) = 2 > 0$, la función no es cóncava.

Ejemplo 5.10. La función $f(x) = \log(x)$, definida sobre $X = (0,\infty)$, es cuasiconvexa, pero no es convexa. Para todo número real t, el contorno inferior de nivel t es el intervalo $(0,e^t]$ y, por lo tanto, la función es cuasiconvexa. Por otro lado, para todo $x \in X$, $f''(x) = -x^{-2} < 0$, por lo que la función no es convexa.

Ejemplo 5.11. La función $f(x,y) = xy$, definida sobre

$$X = \mathrm{R}^2_{++} = \{(x,y) \in \mathrm{R}^2 : x>0,\ y > 0\},$$

es cuasicóncava, pero no es cóncava. Para apreciar que no es cóncava sólo es necesario considerar su matriz hessiana

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

que es indefinida en cada punto del dominio. Para ver que la función es cuasicóncava, consideremos la función $h(x,y) = \log[f(x,y)] = \log(x) + \log(y)$, que es una transformación estrictamente creciente de f y que tiene por matriz hessiana

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

que es definida negativa en cada punto del dominio, por lo que h es (estrictamente) cóncava. Recordemos de pasada que la relación "ser transformación estrictamente creciente de" es simétrica, es decir, el hecho de que h lo es de f implica que f también lo es de h; en nuestro caso, es sencillo comprobar que $f(x,y) = e^{h(x,y)}$.

Ejemplo 5.12. La función $f(x,y) = \log(1 + x^2 + y^2)$, definida sobre $X = \mathrm{R}^2$, es cuasiconvexa, pero no es convexa. Con la indicación de que la función $h(x,y) = 1 + x^2 + y^2$ es convexa, os dejamos que comprobéis el resto de afirmaciones tal como hemos hecho en el último ejemplo.

Dejamos al estudiante...

... la tarea de comprobar la validez de las afirmaciones que aparecen a continuación.

Ejemplo 5.13. Sea $X \subset \mathrm{R}^n$ y $f : X \to \mathrm{R}$ una función cuasicóncava. Definamos una nueva función h sobre el mismo dominio mediante $h(v) = -f(v)$, para cada $v \in X$. Entonces h es una función cuasiconvexa.

Ejemplo 5.14. Sea $X \subset \mathrm{R}^n$ y $f : X \to \mathrm{R}$ una función cuasicóncava. Sea $q : \mathrm{R} \to \mathrm{R}$ una función estrictamente decreciente. Definamos una nueva función $h : X \to \mathrm{R}$ mediante $h(v) = q[f(v)]$ (es decir, h es una transformación estrictamente decreciente de f). Entonces h es una función cuasiconvexa.

Cuando todas las restricciones de un problema normalizado de maximización estén definidas por funciones cuasiconvexas, el conjunto factible siempre será un conjunto convexo. Lo mismo sucederá en un problema de minimización cuando las restricciones estén definidas por funciones cuasicóncavas. En este caso, podemos generalizar el teorema de globalidad que hemos dado al inicio del apartado.

Teorema

Supongamos que el problema normalizado de maximización 5.3 satisface las siguientes hipótesis:

1) La función objetivo f es cóncava, y g_i es una función cuasiconvexa, para $i = 1,2\ldots m$.

2) Todas las funciones son continuamente diferenciables.

Entonces todo punto que satisface las condiciones de Kuhn y Tucker es un maximizador global.

Teorema

Supongamos que el problema normalizado de minimización 5.4 satisface las hipótesis siguientes:

1) La función objetivo f es convexa, y g_i es una función cuasicóncava, para $i = 1,2...m$.

2) Todas las funciones son continuamente diferenciables.

Entonces todo punto que satisface las condiciones de Kuhn y Tucker es un minimizador global.

Cuando queremos que la función objetivo sea cuasicóncava o cuasiconvexa, los resultados se generalizan, a no ser que se trate de un caso excepcional.

Teorema

Supongamos que el problema normalizado de maximización 5.3 satisface las hipótesis siguientes:

1) La función objetivo f es cuasicóncava y g_i es una función cuasiconvexa, para $i = 1, 2\ldots m$.

2) Todas las funciones son continuamente diferenciables.

Entonces todo punto que satisface las condiciones de Kuhn y Tucker es un maximizador global, siempre que no tenga todas las derivadas parciales iguales a cero.

Teorema

Supongamos que el problema normalizado de minimación 5.4 satisface las hipótesis siguientes:

1) La función objetivo f es cuasiconvexa, y g_i es una función cuasicóncava, para $i = 1, 2\ldots m$.

2) Todas las funciones son continuamente diferenciables.

Entonces, todo punto que satisface las condiciones de Kuhn y Tucker es un minimizador global, siempre que no tenga todas las derivadas parciales iguales a cero.

Ejemplo 5.15. Consideremos el problema

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

Normalizando, tenemos:

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

Antes hemos observado que la función objetivo $f(x,y) = xy$ es cuasicóncava sobre el conjunto factible. Por otro lado, las funciones que definen las restricciones (después de normalizar) son todas convexas y, por lo tanto, cuasiconvexas. Las condiciones de Kuhn-Tucker tienen por solución los dos puntos en los que (x, y, λ) valen $(1, 1, 1/2)$ y $(0, 0, 0)$. Sin embargo, $\nabla f(1, 1) = (1, 1) \neq (0, 0)$, pero $\nabla f(0, 0) = (0, 0)$. El teorema anterior nos asegura que el primer punto es un maximizador global, pero no se pronuncia sobre el segundo punto. De hecho, (0, 0) **no** es un maximizador, sino que es un punto de silla de *f*, que aparece aquí porque todo punto crítico (sin restricciones) de una función siempre es solución de las condiciones de KT si está dentro del conjunto factible. Si la función fuera cóncava esto no sería un problema, ya que todo punto en el que todas las derivadas parciales son cero es automáticamente un maximizador global; esto no es necesariamente cierto cuando una función es cuasicóncava, como ilustra este ejemplo.

5.7. Problemas con restricciones de igualdad y de desigualdad

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

En lo que respecta a las condiciones de globalidad en términos de concavidad y convexidad, los teoremas que hemos dado anteriormente continúan siendo aplicables siempre que las restricciones de igualdad sean lineales (lo podemos ver escribiendo una igualdad como dos desigualdades con signos opuestos: $a = b \Leftrightarrow a \leq b$ y $a \geq b$).

Si tenemos restricciones de igualdad que no son lineales, podemos aplicar un argumento de globalidad como el que hemos utilizado en el caso en que sólo hay restricciones de igualdad. Es decir, consideramos para toda solución de las condiciones de KT el lagrangiano después de sustituir el valor óptimo de los multiplicadores, y observamos si la función resultante es cóncava (si maximizamos) o convexa (si minimizamos); si la respuesta es afirmativa, entonces la solución correspondiente es global.

5.8. Resumen

En este módulo, hemos aprendido unas técnicas de optimización que nos permiten resolver problemas no lineales sin restricciones o con restricciones en forma de igualdad o de desigualdad.

También hemos visto condiciones bajo las cuales podemos afirmar que las soluciones que encontramos con nuestras técnicas son de aspecto global. No hemos visto demostraciones matemáticas, pero en cambio sí hemos visto una ilustración de por qué las técnicas funcionan, así como ejemplos de casos en los que no lo hacen.

El énfasis de la presentación ha consistido en dotarnos de técnicas para resolver problemas y de la suficiente intuición para entender cuándo pueden fallar estas técnicas y por qué.

Actividades

1. Analizad el punto crítico (0, 0, 0) de la función $f(x,y,z) = x^3 + 3xy + 3xz + y^3 + 3yz + z^3$ mediante el criterio de los menores principales.

2. Dados los parámetros a y b, sea $f(x, y) = \frac{1}{2}ax^2 - axy + \frac{1}{2}by^2$

a) Probad que (0,0) es un punto crítico de f.
b) Encontrad tres valores del par (a,b) que hacen que (0,0) sea un maximizador, minimizador y punto de silla, respectivamente. (Nota: vuestra respuesta debe ser de la forma: si $a = 12345$ y $b = 8888776$, entonces (0,0) es un maximizador, porque.... donde...).

3. Sea $f : \mathsf{R} \to \mathsf{R}$ definida por $f(x) = 3x^4 - 4x^3$. Usad el hecho de que f es diferenciable para responder a las cuestiones siguientes:
a) ¿En qué intervalos f es creciente y en cuáles es decreciente?
b) ¿En qué intervalos f es cóncava y en cuáles es convexa?

4. Considerad el problema

$$\begin{array}{l} \min\ 2x^2 - 12x + y^2 \\ (x,y) \\ \text{s.a. } y = 2x. \end{array}$$

a) Solucionadlo substituyendo el valor de y en la función objetivo.
b) Solucionadlo mediante el método de Lagrange, y comprobad que los resultados son los mismos.
c) Discutid el carácter global de las soluciones.

5. Encontrad los cuatro puntos críticos de la función $f(x,y) = y - x^2$ sobre la superficie definida por $g(x,y) = x^2 + y^2 = 1$. El conjunto factible es una circunferencia y, por lo tanto, es compacto, pero no es un conjunto convexo. Discutid el carácter global de los puntos críticos.

6. Encontrad los puntos críticos de la función $f(x,y) = x^2 + (y + 1)^2$ sobre la superficie definida por $x^2 + y = 1$.

7. Considerad el problema siguiente:

$$\begin{array}{l} \min\ x^2 + y^2 \\ (x,y) \\ \text{s.a. } (x+1)^3 + y^2 = 0 \end{array}$$

a) Mostrad que el método de Lagrange no da ninguna solución.
b) Sustituid el valor de y^2 que da la restricción en la función objetivo y mostrad que el problema tiene solución.
c) ¿Sabríais decir por qué el método de Lagrange ha fallado?

8. Considerad el problema

$$\begin{array}{l} \min\ x - y \\ (x,y) \\ \text{s.a. } x^2 + y \le 1 \\ \qquad\ \ x \le 0. \end{array}$$

Escribid las condiciones de Kuhn y Tucker, encontrad todos los puntos críticos y después discutid su carácter global.

9. Considerad el problema

$$\begin{array}{l} \max\ -x^2 + y \\ (x,y) \\ \text{s.a. } x^2 + y^2 \le 1 \\ \qquad\ \ y \ge 0. \end{array}$$

Escribid las condiciones de Kuhn y Tucker, encontrad todos los puntos críticos y después discutid su carácter global.

10. Considerad el problema

$$\begin{array}{l} \max\ 2x + y \\ (x,y) \\ \text{s.a. } 2x - x^2 \ge y \end{array}$$

Escribid las condiciones de Kuhn y Tucker, encontrad todos los puntos críticos y después discutid su carácter global.

Ejercicios de autoevaluación

1. Seleccionad la opción correcta. La función…

		x			
		0	1	2	3
	1	0	1	2	3
y	2	0	2	4	6
	3	0	3	6	9

a) tiene un máximo en (3,3).
b) tiene un punto de inflexión en (2,1).
c) tiene un mínimo en (1,2).

2. El rectorado de la Universidad Opulenta de Cataluña ha decidido otorgar una beca de estudios a los estudiantes que tengan una media más alta en su expediente académico (teniendo en cuenta que suspendido = 0; aprobado = 1, etc.). Formulad esto como un problema de optimización, identificando claramente el conjunto factible y la función objetivo. ¿Qué son en este problema el máximo y el maximizador o los maximizadores?

3. De entre todos los triángulos isósceles que tienen perímetro 2, queremos saber cuál es el que tiene un área mayor. Formulad un programa de optimización que nos dé la respuesta, identificando claramente la función objetivo, el conjunto factible, las variables de decisión y las restricciones.

4. Encontrad los puntos críticos de la función $f(x,y)=\frac{1}{3}x^3+x^2+\frac{1}{2}y^2-xy$, y determinad si son maximizadores locales, minimizadores locales o puntos de silla.

5. Encontrad los puntos críticos de la función $f(x,y)=\frac{1}{3}x^3-\frac{1}{2}x^2-\frac{1}{2}y^2-xy-x$, y determinad si son maximizadores locales, minimizadores locales o puntos de silla.

6. Encontrad y analizad los puntos críticos de la función $f(x,y)=2xy-\frac{1}{2}x^4-y^2$.

7. Encontrad y analizad los puntos críticos de la función $f(x,y)=2xy+\frac{1}{2}x^4+y^2$.

8. Aplicad el test de la segunda derivada para determinar si las siguientes funciones son cóncavas, convexas o ninguna de las dos cosas.

a) $f(x) = x^6$
b) $f(x) = \log x$, $(x>0)$
c) $f(x) = e^x$
d) $f(x,y) = 2x-3y$
e) $f(x,y) = xy$
f) $f(x) = 1 + x + e^{-x}$
g) $f(x,y) = x^2 + y^2$
h) $f(x,y) = x^3 - y^3$
i) $f(x,y) = x^{1/3}y^{1/3}$, $(x > 0, y > 0)$
j) $f(x,y) = \sqrt{x} + \sqrt{y}$, $(x \geq 0, y \geq 0)$.

9. Sea $f(x,y)=\frac{1}{4}x^4+\frac{1}{4}y^4+\frac{1}{2}x^2y^2$. Encontrad y analizad los puntos críticos. Discutid su carácter global.

10. Solucionad el problema

$$\begin{array}{l} \min\limits_{(x,y)} x^2 + y^2 \\ \text{s.a. } xy = 1 \end{array}$$

11. Resolved el problema de maximización con restricciones

$$\begin{array}{l} \max\limits_{(x,y)} xy \\ \text{s.a. } 0 \leq x \leq 1 \\ 0 \leq y \leq 1 \end{array}$$

12. Demostrad que el problema

$$\begin{aligned} &\min_{x} e^{x} \\ &\text{s.a. } x > 0 \end{aligned}$$

no tiene ninguna solución y que, en cambio, el problema

$$\begin{aligned} &\min_{x} e^{x} \\ &\text{s.a. } x \geq 0 \end{aligned}$$

sí la tiene.

Solucionario

1. a)

2. Para cada estudiante tendremos un vector cuyos componentes serán todas sus notas. El conjunto factible es el conjunto de todos estos vectores. La función objetivo asocia a cada vector de notas será su media aritmética. El máximo es la nota media más alta, y los maximizadores son todos aquellos estudiantes que tengan esta nota de media.

3. Sea x la mitad de la longitud de la base; y la altura del triángulo y z, la longitud del lado repetido. Entonces el problema que debemos solucionar es

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

Observad que la última igualdad es equivalente a $2x + 2z = 2$. Para solucionar el problema debemos escribir el lagrangiano:

$$L(x, y, z, \lambda_1, \lambda_2) = xy - \lambda_1(x^2 + y^2 - z^2) - \lambda_2 (x + z - 1).$$

Las condiciones de primer orden son

$$\begin{aligned} \frac{\partial L}{\partial x} &= y - 2\lambda_1 x - \mu = 0 \\ \frac{\partial L}{\partial x} &= x - 2\lambda_1 y = 0 \\ \frac{\partial L}{\partial x} &= 2\lambda_1 z - \mu = 0 \\ x^2 + y^2 - z^2 &= 0 \\ x + z &= 1 \end{aligned}$$

Operando algebraicamente, encontramos la solución

$$(x, y, z, \lambda_1, \lambda_2) = (1/3, 1/\sqrt{3}, 2/3, 1/2\sqrt{3}, 2/3\sqrt{3})$$

Observad que la solución es un triángulo equilátero, ya que $z = 2x$.

4. Encontramos los puntos críticos igualando las derivadas parciales a cero:

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

A partir de la segunda ecuación tenemos que $y = x$, y sustituyendo en la primera, encontramos que las soluciones son $(x,y) = (0,0)$ y $(x,y) = (-1,-1)$. La matriz hessiana es

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

Sustituyendo $(x,y) = (0,0)$, tenemos que:

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

Dado que esta matriz es definida positiva, deducimos que $(0,0)$ es un minimizador local de f.

Si sustituimos $(x,y) = (-1, -1)$, tenemos que:

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

Como esta matriz es indefinida, deducimos que (–1,–1) es un punto de silla de f.

5. Encontramos los puntos críticos igualando las derivadas parciales a cero:

$$\frac{\partial f}{\partial x} = x^2 - x - y - 1 = 0$$
$$\frac{\partial f}{\partial x} = \quad -y - x = 0$$

A partir de la segunda ecuación, tenemos que $y = -x$, y sustituyendo la primera encontramos que las soluciones son $(x, y) = (1, -1)$ y $(x, y) = (-1, 1)$. La matriz hessiana es

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

Sustituyendo $(x, y) = (1, -1)$, tenemos que:

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

Dado que esta matriz es indefinida positiva, deducimos que $(1, -1)$ es un punto de silla de f. Si sustituimos $(x, y) = (-1, 1)$, tenemos que:

$$D^2 f(-1, 1) = \begin{pmatrix} -3 & -1 \\ -1 & -1 \end{pmatrix}$$

Dado que esta matriz es definida negativa, deducimos que $(-1, 1)$ es un maximizador local de f.

6. Encontramos los puntos críticos igualando las derivadas parciales a cero:

$$\frac{\partial f}{\partial x} = 2y - 2x^3 = 0$$
$$\frac{\partial f}{\partial y} = 2x - 2y = 0$$

A partir de la segunda ecuación tenemos que $y = x$, y sustituyendo en la primera encontramos que las soluciones son $(x, y) = (0, 0)$, $(x, y) = (1, 1)$ y $(x, y) = (-1, -1)$. La matriz hessiana es

$$D^2 f(x, y) = \begin{pmatrix} -6x^2 & 2 \\ 2 & -2 \end{pmatrix}$$

Sustituyendo $(x, y) = (0, 0)$, tenemos que:

$$D^2 f(0, 0) = \begin{pmatrix} 0 & 2 \\ 2 & -2 \end{pmatrix}$$

Dado que esta matriz es indefinida, deducimos que $(0, 0)$ es punto de silla de f.

Si sustituimos otros valores de (x, y), tenemos que:

$$D^2 f(1, 1) = D^2 f(-1, -1) = \begin{pmatrix} -6 & 2 \\ 2 & -2 \end{pmatrix}$$

Dado que esta matriz es indefinida, deducimos que $(1, 1)$ y $(-1, -1)$ son puntos de silla de f.

7. Encontramos los puntos críticos igualando las derivadas parciales a cero:

$$\frac{\partial f}{\partial x} = 2y + 2x^3 = 0$$
$$\frac{\partial f}{\partial y} = 2x + 2y = 0$$

A partir de la segunda ecuación tenemos que $y = -x$, y sustituyendo en la primera encontramos que las soluciones son $(x, y) = (0, 0)$, $(x,y) = (1, -1)$ y $(x, y) = (-1, 1)$. La matriz hessiana es

$$D^2f(x, y) = \begin{pmatrix} 6x^2 & 2 \\ 2 & 2 \end{pmatrix}$$

Sustituyendo $(x, y) = (0, 0)$, tenemos que:

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

Como esta matriz es indefinida, deducimos que (0,0) es un punto de silla de f.

Si sustituimos los otros valores de (x,y), tenemos que:

$$D^2f(1, 1) = D^2f(-1, 1) = \begin{pmatrix} 6 & 2 \\ 2 & 2 \end{pmatrix}$$

Dado que esta matriz es definida positiva, deducimos que (1, –1) y (–1, 1) son minimizadores locales de f.

8.

a) $f''(x) = 12x^4 \Rightarrow$ convexa

b) $f''(x) = \dfrac{-1}{x^2} \Rightarrow$ cóncava

c) $f''(x) = e^x \Rightarrow$ convexa

d) $D^2f(x, y) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow$ cóncava y convexa

e) $D^2f(x, y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Rightarrow$ ni cóncava ni convexa

f) $f''(x) = e^{-x} \Rightarrow$ convexa

g) $D^2f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \Rightarrow$ convexa

h) $D^2f(x, y) = \begin{pmatrix} 6x & 0 \\ 0 & -6y \end{pmatrix} \Rightarrow$ ni cóncava ni convexa

i) $D^2f(x, y) = \dfrac{2}{9}x^{-2/3}y^{-2/3}\begin{pmatrix} -2\dfrac{y}{x} & 1 \\ 1 & -2\dfrac{x}{y} \end{pmatrix} \Rightarrow$ cóncava

j) $D^2f(x, y) = \begin{pmatrix} -\dfrac{1}{4}x^{-3/4} & 0 \\ 0 & -\dfrac{1}{4}y^{-3/4} \end{pmatrix} \Rightarrow$ cóncava

9. Encontremos los puntos críticos igualando las derivadas parciales a cero:

$$\frac{\partial f}{\partial x} = x^3 + xy^2 = 0$$
$$\frac{\partial f}{\partial y} = y^3 + x^2y = 0$$

La única solución es $(x, y) = (0, 0)$. La matriz hessiana es

$$D^2f(x, y) = \begin{pmatrix} 3x^2 + y^2 & 2xy \\ 2xy & 3y^2 + x^2 \end{pmatrix}$$

Sustituyendo $(x, y) = (0, 0)$ tenemos que:

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

El criterio local no es decisorio.

Sin embargo, cualquiera que sea (x,y), podemos ver que $D^2f(x,y)$ es una matriz semidefinida positiva, y esto implica que la función es convexa. Esto comporta también que (0,0), al ser un punto crítico de una función convexa, sea un minimizador global.

10. El lagrangiano es $L(x,y,\lambda) = x^2 + y^2 - \lambda(xy - 1)$, y las condiciones de primer orden son:

$$\frac{\partial L}{\partial x} = 2x - \lambda y = 0$$
$$\frac{\partial L}{\partial y} = 2y + (-\lambda x = 0)$$
$$xy - 1 = 0$$

Las soluciones son $(x, y, \lambda) = (1, 1, 2)$ y $(x, y, \lambda) = (-1, -1, 2)$.

11. Escribimos el problema de forma normalizada:

$$\max_{(x,y)} xy$$
$$\text{s.a.} \quad -x \le 0$$
$$-y \le 0$$
$$x - 1 \le 0$$
$$y - 1 \le 0$$

El lagrangiano es

$$L(x,y,\lambda_1,\lambda_2,\lambda_3,\lambda_4) = xy + \lambda_1 x + \lambda_2 y - \lambda_3(x-1) - \lambda_4(y-1).$$

Las condiciones de Kuhn y Tucker son

$$\frac{\partial L}{\partial x} = y + \lambda_1 - \lambda_3 = 0$$
$$\frac{\partial L}{\partial y} = x + \lambda_2 - \lambda_4 = 0$$
$$\lambda_1 \ge 0, x \ge 0, \lambda_1 x = 0$$
$$\lambda_2 \ge 0, y \ge 0, \lambda_2 y = 0$$
$$\lambda_3 \ge 0, x \le 1, \lambda_3(x-1) = 0$$
$$\lambda_4 \ge 0, y \le 1, \lambda_4(y-1) = 0$$

Las soluciones son

$$(x,y,\lambda_1,\lambda_2,\lambda_3,\lambda_4) = (0,0,0,0,0,0),$$
$$(x,y,\lambda_1,\lambda_2,\lambda_3,\lambda_4) = (1,1,0,0,1,1).$$

La primera solución corresponde a un punto de silla de la función objetivo, y la podemos descartar. La segunda solución es el maximizador del problema.

1.2. El problema

$$\min_x e^x$$
$$\text{s.a. } x > 0$$

es un problema de optimización sin restricciones, ya que el conjunto factible está definido por una desigualdad estricta y, por lo tanto, es abierto. Si hay alguna solución, entonces la derivada de la función objetivo en este punto debe ser 0. Pero si $f(x) = e^x$, entonces $f'(x) = e^x > 0$, para todo x. Por lo tanto, no puede haber ninguna solución.

El problema

$$\min_x e^x$$
$$s.a. \; x \ge 0$$

es un problema de optimización con restricciones de desigualdad. El lagrangiano es

$$L(x, \lambda) = e^x - \lambda x.$$

Las condiciones de Kuhn-Tucker son:

$$\frac{\partial L}{\partial x} = e^x - \lambda = 0$$
$$\lambda \geq 0,\ x \geq 0,\ \lambda x = 0.$$

Tenemos

$$\lambda = e^x > 0 \Rightarrow x = 0$$

y ésta es la única solución. La restricción es lineal y la función objetivo es convexa, así que la solución es global.

Glosario

condiciones de exclusión
Condiciones que se cumplen cuando se da una restricción de desigualdad débil y esta restricción se satisface como desigualdad estricta; el multiplicador correspondiente, entonces, es cero.

condiciones de Kuhn y Tucker
Condiciones necesarias para la optimización en términos de las derivadas de primer orden.

condiciones de segundo orden
Condiciones para la optimización en términos de las derivadas de segundo orden.

condiciones de tangencia: Condiciones que, en problemas con restricciones, expresan que la curva de nivel correspondiente al valor óptimo de la función objetivo no puede cortar el conjunto factible.

función objetivo
Función que queremos maximizar o minimizar.

lagrangiano
Función objetivo modificada, en la que introducimos un término para toda restricción, ponderado por el modificador correspondiente.

máximo:
Mayor valor alcanzado por la función objetivo.

máximo global
Valor máximo de la función objetivo, pero sólo en un cierto entorno.

maximizador
Punto del dominio en el que la función alcanza el valor máximo. Puede haber muchos maximizadores, pero sólo hay un máximo.

mínimo
Menor valor alcanzado por la función objetivo.

mínimo global
Valor mínimo de la función objetivo sobre todo el dominio en consideración (teniendo en cuenta las posibles restricciones).

mínimo local
Valor mínimo de la función objetivo, pero sólo en un cierto entorno.

minimizador
Punto del dominio en el que la función alcanza el valor mínimo. Puede haber muchos minimizadores, pero sólo hay un mínimo.

punto crítico
Candidato a maximizador o minimizador. Se trata de un punto que satisface las condiciones de primer orden de un problema de optimización.

punto de silla
Punto en el que todas las derivadas parciales son cero, y a partir del cual la función crece en alguna dirección y decrece en alguna otra.

regularidad
Independencia lineal de los gradientes de las restricciones que se satisfacen con igualdad en un punto.

restricciones
Condiciones, en forma de igualdad o de desigualdad que deben cumplir las variables en un problema de optimización.

Bibliografía

Encontraréis una presentación clara y concisa de lo que hemos tratado en este módulo en el libro siguiente:

Borrell, Josep (1990). *Métodos matemáticos para la economía. Programación matemática*. Madrid: Ediciones Pirámide.

Para una exposición más completa y a un nivel más avanzado, podéis consultar:

Chiang, Alpha (1987). *Métodos fundamentales de la economía matemática*. McGraw-Hill (México, 1990).

Luenberger, David (1992). *Programación lineal y no lineal*. Wilmington: Addison-Wesley Iberoamericana.

Un libro también más avanzado, pero con buenas explicaciones intuitivas y gráficas, y dirigido a economistas, es

Dixit, Avinash. *Optimization in Economic Theory*. Oxford.

Finalmente, lo que hacemos aquí y mucho más está bien explicado en:

Sydsaeter, Knut; Hammond, Peter (1995). *Mathematics for Economic Analysis*. Englewood Cliffs, N.J.: Prentice-Hall International.