Universidad Autónoma del Estado de México

Facultad de Ciencias
Licenciatura en Matemáticas

Desigualdades Restringidas y el Teorema de Kuhn y Tucker

TESIS

QUE PARA OBTENER EL TÍTULO DE:
Matemático

PRESENTA:
Arisbeth González González

DIRECTOR DE TESIS:
Dr. Alfredo Cano Rodríguez

2021

Introducción

Los problemas del mundo real son, en general, no lineales, mixtos y multivariables. Poder resolver esa clase de problemas es importante en muchos de los casos encontrar la solución óptima, esto suele ser el principal objetivo, entenderemos por valor óptimo al valor "ideal" dependiendo de la naturaleza del problema, aunque muchas veces es suficiente con encontrar soluciones con ciertas características, aunque se trate de un resultado sub-óptimo pero que determina cierta "calidad" de esta solución respecto del valor ideal.

Las técnicas de optimización se enfocan en determinar el conjunto de restricciones a seguir para maximizar o minimizar algunos tipos de funciones. Dicha respuesta, en general, es un indicador del tipo "Costo", "Producción", "Ganancia", etc., la cual es una función de las restricciones seleccionadas. Dicha respuesta se denomina objetivo, y la función asociada se llama función objetivo. Un conjunto de restricciones es un determinado conjunto de valores que toman los factores que intervienen en el problema y que podemos controlar a fin de regular el rendimiento de la función objetivo, es decir, en lenguaje matemático son las variables independientes de la función.

En general, la dificultad en los problemas de optimización radica en que estamos limitados en nuestro poder de decisión. Por ejemplo, podemos tener un tope máximo a la cantidad de horas trabajadas si el problema es de tipo económico, al igual que un rango de cantidad de operarios entre los que podemos manejar. Estas limitaciones, llamadas "restricciones", reducen la cantidad de alternativas posibles, definiendo un espacio acotado de soluciones factibles (y complicando, de paso, la resolución del problema). El término "soluciones factibles" hace referencia a las posibles soluciones del problema, dentro de las que dan resultados reales al problema estudiado. Pueden existir una o más soluciones óptimas, es decir, aquellas que, además de cumplir con todas las restricciones, maximizan (o minimizan, según sea el problema a resolver) el valor de la función objetivo.

Resumiendo, un problema de optimización está compuesto de los siguientes elementos:

1. Un conjunto de restricciones.
2. Un conjunto de soluciones factibles, el cual contiene todas las posibles combinaciones de valores de variables independientes que satisfacen las restricciones anteriores.
3. Una función objetivo, que vincula las soluciones factibles con la estructura del problema.

Dentro de la práctica de la optimización, existen dos ramas relacionadas pero diferenciadas entre sí.

Por un lado, nos encontramos con el problema de cómo modelar adecuadamente el sistema de estudio, y por otro cómo resolver el modelo. En general, la matemática se ocupa de buscar mejores métodos de solución, mientras que la ingeniería analizaformas de modelar sistemas reales.

En base a su naturaleza, hay varias formas de clasificar un problema de optimización. Analizar en qué categoría entra es importante para definir el método de solución a utilizar, ya que no hay un método único.

Para comenzar, una primera distinción la podemos realizar en base a la continuidad o no de las variables de decisión. Se dice que estamos frente a un problema de “Optimización Continua” cuando todas las variables de decisión pueden tomar cualquier valor perteneciente al conjunto de los números reales, dentro de este tipo de problemas, son de particular importancia los problemas de “Optimización Convexa", en los cuales se debe minimizar una función convexa sujeta a un conjunto solución convexo.

Cuando tanto la función objetivo como las restricciones son lineales, hablamos de un problema de “Optimización Convexa Lineal” o "Programación Lineal". En el caso de trabajar con variables discretas (es decir, que solo puedan tomar valores enteros) nos enfrentamos a un problema de “Optimización Combinatoria”. Por raro que pueda parecer, en general un problema de optimización combinatoria es más complicado de resolver que uno de optimización continua. En el medio tenemos los problemas de “Optimización

Mixta” en los cuales algunas variables son continuas y otras son discretas.

Los métodos de resolución de problemas de optimización se pueden clasificar en tres tipos diferentes:

1. Resolución mediante cálculo.
2. Resolución mediante técnicas de búsquedas.
3. Resolución mediante técnicas de convergencia de soluciones.

Los métodos de resolución por cálculo emplean derivadas para determinar para qué valores del dominio la función presenta un máximo o un mínimo, dentro de estos métodos tenemos té cnicas para abarcar una gran variedad de problemas.

Dedicaremos este trabajo a mostrar algunos métodos de solución a problemas de optimización, haremos el análisis en términos de problemas matemáticos y mostraremos algunos métodos de solución. En el Capítulo 1 daremos conceptos preliminares y una serie de resultados que serán empleados en los capítulos posteriores.

En el Capítulo 2 analizaremos el Teorema de Lagrange, una herramienta del cálculo de varias variables para encontrar extremos condicionados del tipo $g(x) = 0$, donde $g(x)$ representa el conjunto de restricciones, definimos la función de Lagrange, el vector de multiplicadores de Lagrange, daremos una demostración del Teorema de Lagrange, y mostraremos con algunos ejemplos la aplicación de tal teorema, a este proceso se le suele llamar método de los multiplicadores de Lagrange.

En el Capítulo 3 enunciamos el Teorema de Kuhn y Tucker, el cual proporciona una elegante caracterización del comportamiento de la función objetivo f y las funciones de restricción h_i en el óptimo local de problemas de optimización restringidos por la desigualdad $h_i(x) \geq 0$. Las condiciones que describe se pueden ver como las condiciones necesarias de primer orden para obtener el óptimo local en este tipo de problemas.

Al igual que la condición análoga del teorema de Lagrange la condición del Teorema de Kuhn y Tucker de que el rango de la matriz de derivada de las restricciones debe ser máximo, se denomina cualificación de restricción. Esta condición juega un papel central en la prueba del teorema. Además si falla la cualificación de la restricción, el teorema mismo podría fallar y lo ilustramos en este punto mediante un ejemplo.

El vector λ en el Teorema de Kuhn y Tucker se llama vector multiplicador de Kuhn y Tucker correspondiente al máximo local x^*. Al igual que con los multiplicadores de Lagrange, también se puede pensar que los multiplicadores de Kuhn y Tucker miden la sensibilidad de la función objetivo en x para las relajaciones de las diversas limitaciones. De la misma manera que en el capítulo anterior daremos una demostración del teorema y mostraremos el procedimiento para aplicarlo en problemas de óptimos locales.

Finalmente presentamos dos ejemplos extraídos de economía, que ilustran el uso del procedimiento para encontrar soluciones a problemas de optimización restringidos por desigualdad. El primer ejemplo, considera un problema de maximización de la utilidad. El segundo ejemplo describe un problema de minimización de costos.

Finalmente en el Capítulo 4 desarrollamos algunos aspectos de la Optimización Convexa, la cual es una rama de las técnicas de optimización que trata sobre técnicas de minimización de funciones convexas sobre un dominio también convexo, a esta técnica se le llama Programación Convexa (PC). El conjunto convexo a menudo se define en términos de desigualdades que involucran otras funciones convexas. Comenzaremos describiendo los problemas básicos. Luego discutimos las caracterizaciones de las soluciones dadas por el Teorema de Karush-Kuhn-Tucker (KKT), mostraremos dos versiones del Teorema.

Índice general

Capítulo 1

Preliminares.

Consideramos D un subconjunto de $\mathbb{R}^n$. Una función escalar de n variables reales es una regla de correspondencia que a cada punto $\overline{x} \in D$, le asigna un único número real $f(\overline{x})$. La función f se representa

$$f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}.$$

El conjunto D sobre el que está definida la función f se llama dominio de f y, dado $\overline{x} \in D$, el número $f(\overline{x})$ se llama imagen de $\overline{x}$ por f.

Consideremos un subconjunto D de $\mathbb{R}^n$. Una función vectorial de n variables y m coordenadas es una regla de correspondencia que a cada elemento del dominio D le asigna un único vector de m números reales.

Para indicar lo anterior se suele escribir

$$f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m.$$

La norma de un vector $\overline{x} \in \mathbb{R}^n$ se define como el escalar

$$\| \overline{x} \| = \sqrt{\overline{x} \cdot \overline{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

este valor geométricamente se trata de la longitud del vector $\overline{x}$.

La distancia entre dos puntos de $\mathbb{R}^n$ se define como

$$d(\overline{x}, \overline{y}) = \| \overline{x} - \overline{y} \| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}.$$

Dado $\epsilon > 0$ y $\overline{p} \in \mathbb{R}^n$, se define la bola abierta de centro $\overline{p}$ y radio ε como el conjunto

$$B_\varepsilon(\overline{p}) = \{\overline{x} \in \mathbb{R}^n : \| \overline{x} - \overline{p} \| < \varepsilon\}.$$

Un subconjunto A de $\mathbb{R}^n$ es abierto si para todo punto $\overline{p} \in A$ existe un $\varepsilon > 0$ tal que $B_\varepsilon(\overline{p}) \subset A$, en otras palabras, A es abierto si cuando estamos en un punto $\overline{p}$ de A podemos movernos en cualquier dirección sin salirnos de A.

Un subconjunto C de $\mathbb{R}^n$ es cerrado si su complemento $\mathbb{R}^n \setminus C$ es abierto.

Un punto $\overline{p} \in \mathbb{R}^n$ es un punto de acumulación de un subconjunto A de $\mathbb{R}^n$ si para todo $\varepsilon > 0$ existe un punto $\overline{x} \in A$, $\overline{x} \neq \overline{p}$, tal que $\| \overline{x} - \overline{p} \| < \varepsilon$. En otros términos, $\overline{p}$ es un punto de acumulación de A si hay puntos de A distintos del propio $\overline{p}$ tan cercanos a $\overline{p}$ como se desee.

Si $\overline{p} \notin A$ no es un punto de acumulación de A, se dice que $\overline{x} \in A$ es un punto aislado de A, es decir, un punto de A es aislado si no tiene a su alrededor ningún otro punto de A.

Sea $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función escalar y sea $\overline{p} \in \mathbb{R}^n$ un punto de acumulación de D. Se dice que un punto $l \in \mathbb{R}$ es el límite de $f(\overline{x})$ cuando $\overline{x}$ tiende a $\overline{p}$ (y se representa por $\lim\limits_{\overline{x} \to \overline{p}} f(\overline{x}) = l$) si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que si $\overline{x} \in D$, $\overline{x} \neq \overline{p}$ y $\| \overline{x} - \overline{p} \| < \delta$, entonces $| f(\overline{x}) - l | < \varepsilon$.

Observaciones:

1. Si una función tiene límite en un punto, el límite es único, es decir, una función no puede tender a dos límites distintos en el mismo punto.

2. Una función vectorial $f(\overline{x}) = (f_1(\overline{x}), \ldots, f_m(\overline{x}))$ tiene límite en un punto $\overline{p}$ si y sólo si lo tienen todas las funciones coordenadas f_i, y entonces

$$\lim_{\overline{x} \to \overline{p}} f(\overline{x}) = \left(\lim_{\overline{x} \to \overline{p}} f_1(\overline{x}), \ldots, \lim_{\overline{x} \to \overline{p}} f_m(\overline{x}) \right).$$

***Definición* 1.1.** *Diremos que una función $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ es continua en un punto $\overline{p} \in D$ si el límite de $f(\overline{x})$ cuando $\overline{x}$ tiende a $\overline{p}$ es igual a $f(\overline{p})$.*

Una función vectorial $f(\overline{x}) = (f_1(\overline{x}), \ldots, f_m(\overline{x}))$ es continua en un punto $\overline{p}$ si y sólo si lo son todas sus funciones coordenadas f_i.

Si $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ es una función escalar definida en un abierto D y $\overline{x} \in D$, llamaremos **función de incrementos parciales** de f respecto de la variable x_i a la función

$$\triangle_{x_i} f(\overline{x}) = f(x_1, \ldots, x_i + \triangle x_i, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n).$$

La función de incrementos parciales de una función f respecto a x_i es otra función cuyas variables son las de f más la nueva variable $\triangle x_i$, y nos permite calcular el incremento que experimenta f cuando la variable x_i se incrementa en la cantidad $\triangle x_i$. Si fijamos al punto $\overline{x}$, obtenemos una función cuya única variable es $\triangle x_i$.

Definición* 1.2.** *Sea $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función escalar definida en un abierto D. Para cada punto $\bar{p} \in D$, definimos la* ***derivada parcial *de f respecto de la variable x_i en el punto $\overline{p}$ como:*

$$\begin{aligned} \left.\frac{\partial f}{\partial x_i}\right|_{\overline{p}} &= lim_{\triangle x_i \longrightarrow 0} \frac{\triangle_{x_i} f(\overline{p})}{\triangle x_i} \\ &= lim_{\triangle x_i \longrightarrow 0} \frac{f(p_1, \ldots, p_i + \triangle x_i, \ldots, p_n) - f(p_1, \ldots, p_n)}{\triangle x_i}, \end{aligned}$$

donde $\triangle_{x_i}$ es el incremento.

Observemos que una derivada parcial es un límite y por lo tanto no tiene por que existir. En caso de que este exista, si $\triangle x_i$ es suficientemente pequeño, entonces

$$\frac{\triangle_{x_i} f(\overline{p})}{\triangle x_i} \approx \left.\frac{\partial f}{\partial x_i}\right|_{\overline{p}}.$$

***Definición* 1.3.** *Sea $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una función vectorial definida en un abierto D. Dado $\overline{p} \in D$, decimos que f es derivable en $\overline{p}$ si existen las derivadas parciales respecto de todas las variables de sus funciones coordenadas, es decir, existen*

$$\frac{\partial f_j}{\partial x_i}(\overline{p}), \text{ para } \quad i = 1, \ldots, n \quad j = 1, \ldots, m,$$

y podemos denotar

$$\frac{\partial f}{\partial x_i}(\overline{p}) = \left(\frac{\partial f_1}{\partial x_i}(\overline{p}), \ldots, \frac{\partial f_m}{\partial x_i}(\overline{p}) \right).$$

Si $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ es una función derivable respecto de x_i en todos los puntos del abierto D, entonces indicamos la función derivada parcial respecto de x_i como

$$\frac{\partial f}{\partial x_i} : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}.$$

Una función $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ definida en un abierto D es de clase C^n en D, donde $n \geq 1$ es un número natural, si existen las derivadas parciales de f hasta orden n y todas ellas son continuas en D. Si f tiene derivadas parciales de todos los órdenes posibles y todas ellas son continuas, entonces se dice que f es de clase C^∞ en D.

Definición* 1.4.** *Si $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ es una función escalar derivable en el abierto D, definimos y denotamos su vector* ***gradiente *como el vector formado por sus derivadas parciales:*

$$\nabla f(\overline{x}) = \left(\frac{\partial f}{\partial x_1}(\overline{x}), \ldots, \frac{\partial f}{\partial x_n}(\overline{x}) \right).$$

Definición* 1.5.** *Si $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es una función vectorial derivable en el abierto D se define la* ***Matriz Jacobiana *de f como la matriz formada por todas las derivadas parciales primeras de sus funciones coordenadas $f_1, \ldots, f_m$ del modo siguiente:*

$$J_f(\overline{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\overline{x}) & \frac{\partial f_1}{\partial x_2}(\overline{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\overline{x}) \\ \frac{\partial f_2}{\partial x_1}(\overline{x}) & \frac{\partial f_2}{\partial x_2}(\overline{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\overline{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\overline{x}) & \frac{\partial f_m}{\partial x_2}(\overline{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\overline{x}) \end{pmatrix}.$$

Definición* 1.6.** *Si $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ es una función escalar dos veces derivable respecto de todas las variables en el abierto D, definimos su* ***Matriz Hessiana *como la matriz formada por sus derivadas parciales de orden* 2 *del modo siguiente:*

$$H_f(\overline{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\overline{x}) & \frac{\partial^2 f}{\partial x_1 x_2}(\overline{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 x_n}(\overline{x}) \\ \frac{\partial^2 f}{\partial x_2 x_1}(\overline{x}) & \frac{\partial^2 f}{\partial x_2^2}(\overline{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 x_n}(\overline{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1}(\overline{x}) & \frac{\partial^2 f}{\partial x_n x_2}(\overline{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^n}(\overline{x}) \end{pmatrix}.$$

Ejemplo. *Sea $f(x,y,z) = x^3y + y^2z + xyz^3$, calculando las derivadas parciales que forman el gradiente de f se tiene:*

$$\bigtriangledown f(x,y,z) = (3x^2y + yz^3, x^3 + 2yz + xz^3, y^2 + 3xyz^2).$$

Por la definición 5 podemos calcular la matriz Jacobiana con respecto al gradiente de la función,

$$J_f(x,y,z) = \begin{pmatrix} 6xy & 3x^2+z^3 & 3yz^2 \\ 3x^2+z^3 & 2z & 2y+3xz^2 \\ 3yz^2 & 2y+3xz^2 & 6xyz \end{pmatrix},$$

mientras que la matriz Hessiana de la función es:

$$H_f(x,y,z) = \begin{pmatrix} 6xy & 3x^2+z^3 & 3yz^2 \\ 3x^2+z^3 & 2z & 2y+3xz^2 \\ 3yz^2 & 2y+3xz^2 & 6xyz \end{pmatrix}.$$

Cada matriz cuadrada A tiene asociado un número real llamado determinante de A, que representaremos por $|A|$ o $detA$.

***Definición* 1.7.** *Una matriz cuadrada es **regular** si su determinante es no nulo y es **singular** si su determinante es nulo.*

*Los **menores principales** de orden k de una matriz $n \times n$ (con $k \leq n$) son los determinantes de las matrices formadas por k filas de A (en orden) y las mismas k columnas.*

*El **menor principal conducente** de orden k es el menor principal formado con las k primeras filas y las k primeras columnas de la matriz.*

Definición* 1.8.** *Si A es una matriz simétrica $n \times n$ de números reales, se llama* ***forma cuadrática *determinada por A a la aplicación $q : \mathbb{R}^n \longrightarrow \mathbb{R}$ dada por $q(\bar{x}) = \bar{x}A\bar{x}^t$.*

***Teorema* 1.1.** *(Criterio de Jacobi) Sea $q : \mathbb{R}^n \longrightarrow \mathbb{R}$ una forma cuadrática cuya matriz asociada es A. Entonces*

1. *En el caso en que A sea regular:*

   a *Si los menores principales conducentes son todos > 0 entonces q es definida positiva.*

   b *Si los menores principales conducentes tienen signos alternados, es decir, si A_i es el menor conducente de orden i*

   $$A_i < 0 \text{ si } i \text{ es impar y } A_i > 0 \text{ si } i \text{ es par,}$$

   entonces q es definida negativa. Nótese que el primero debe ser negativo.

   c *En otro caso q es indefinida.*

2. *En el caso en que A sea singular:*

   a *Si todos los menores principales son: $A_i \geq 0$ entonces A es semidefinida positiva.*

   b *Si los menores principales son: $A_i \leq 0$ si i impar y $A_i \geq 0$ si i par entonces A es semidefinida negativa.*

   c *En otro caso q es indefinida.*

Para aplicar este criterio, en primer lugar hemos de calcular el determinante de A para saber si es regular o singular. Si es regular sólo necesitaremos calcular los menores principales conducentes, mientras que si es singular tendremos que estudiar el signo de todos los menores principales, conducentes o no.

Para su demostración ver [16].

Sea $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función escalar definida en un abierto D y sea $\bar{x}^* \in D$.

- Se dice que $\bar{x}^*$ es un máximo global estricto absoluto de f si para todo $\bar{x} \in D$, $\bar{x} \neq \bar{x}^*$, se cumple $f(\bar{x}^*) > f(\bar{x})$.
- Se dice que $\bar{x}^*$ es un mínimo global estricto absoluto de f si para todo $\bar{x} \in D$, $\bar{x} \neq \bar{x}^*$, se cumple $f(\bar{x}^*) < f(\bar{x})$.

Si $S \subset D$ es un subconjunto arbitrario entonces:

- Se dice que $\bar{x}^*$ es un máximo global estricto relativo a S de f si $\bar{x}^* \in S$ y para todo $V \in D \cap S$, $\bar{x} \neq \bar{x}^*$, se cumple $f(\bar{x}^*) > f(\bar{x})$.
- Se dice que $\bar{x}^*$ es un mínimo global estricto relativo a S de f si $\bar{x}^* \in S$ y para todo $\bar{x} \in D \cap S$, $\bar{x} \neq \bar{x}^*$, se cumple $f(\bar{x}^*) < f(\bar{x})$.
- Se dice que $\bar{x}^*$ es un máximo local estricto relativo a S de f si $\bar{x}^* \in S$ y existe un $\varepsilon > 0$ tal que para todo $\bar{x} \in D \cap S$, $\bar{x} \neq \bar{x}^*$, y $\| \bar{x} - \bar{x}^* \| < \varepsilon$ se cumple $f(\bar{x}^*) > f(\bar{x})$.
- Se dice que $\bar{x}^*$ es un mínimo local estricto relativo a S de f si $\bar{x}^* \in S$ y existe un $\varepsilon > 0$ tal que para todo $\bar{x} \in D \cap S$, $\bar{x} \neq \bar{x}^*$, y $\| \bar{x} - \bar{x}^* \| < \varepsilon$ se cumple $f(\bar{x}^*) < f(\bar{x})$.

Ejemplo. *Consideremos la función dada en la siguiente gráfica*

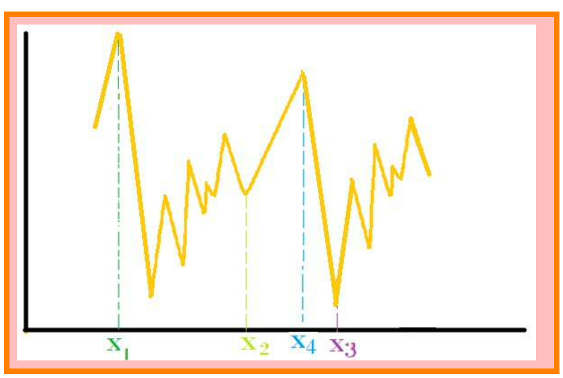

Figura 1.1: Máximos y Mínimos

El punto x_1 es un máximo global absoluto. En este punto la función vale más (por eso es máximo) que en cualquier otro punto del dominio (por eso es global). Es estricto porque no hay otro punto en el dominio que tome su misma imagen.

El punto x_2 es un mínimo global no estricto relativo. Es un mínimo porque en este punto la función vale menos que en cualquier otro punto del dominio, no es estricto pues hay otro punto en el dominio que tome su misma imagen.

El punto x_3 es un mínimo global absoluto. En él la función toma un valor más bajo (mínimo) que en cualquier otro punto del dominio (global). Es estricto porque no hay otro punto en el dominio que tome su misma imagen.

El punto x_4 es un máximo local estricto relativo. Es un máximo porque en este punto la función vale más que en cualquier otro punto del dominio cercano a él y es local porque hay otros puntos donde la función es mayor. Es estricto porque alrededor de él no hay otro punto que tome su misma imagen. A estos puntos les llamaremos en general extremos.

Las siguientes definiciones y teoremas serán utilizadas en los siguientes capítulos.

***Definición* 1.9.** *Un punto crítico de una función $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ de clase C^1 es un punto $\bar{x}^*$ de su dominio tal que $\nabla f(\bar{x}^*) = \bar{0}$.*

***Teorema* 1.2.** *Sea $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función de clase C^1 en un abierto D. Si $\bar{x}^* \in D$ un extremo local de f, entonces $\nabla f(\bar{x}^*) = \bar{0}$.*

Para su demostración ver [16].

***Teorema* 1.3.** *Sea $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función de clase C^2 en el abierto D y sea $\bar{x}^* \in D$ un punto crítico.*

Entonces:

1. *Si $H_f(\bar{x}^*)$ es definida positiva, $\bar{x}^*$ es un mínimo local de f.*
2. *Si $H_f(\bar{x}^*)$ es definida negativa, $\bar{x}^*$ es un máximo local de f.*
3. *Si $H_f(\bar{x}^*)$ es indefinida, $\bar{x}^*$ es un punto de silla de f.*

A este teorema se le conoce como la condición suficiente de segundo orden para extremos condicionados.

Para su demostración ver [16].

Definición* 1.10.** *Sea $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función, f es* ***convexa *si cuando $\overline{x}$, $\overline{y} \in D$ y $0 \leq \lambda \leq 1$, se cumple que*

$$f(\lambda\overline{x} + (1-\lambda)\overline{y}) \leq \lambda f(\overline{x}) + (1-\lambda)f(\overline{y}).$$

Definición* 1.11.** *Se dice que f es* ***estrictamente convexa *si se cumple*

$$f(\lambda\overline{x} + (1-\lambda)\overline{y}) < \lambda f(\overline{x}) + (1-\lambda)f(\overline{y})$$

para todo $\overline{x}$, $\overline{y} \in D$ con $\overline{x} \neq \overline{y}$ y $0 < \lambda < 1$.

Definición* 1.12.** *Sea $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función, f es* ***cóncava *si cuando $\overline{x}$, $\overline{y} \in D$ y $0 \leq \lambda \leq 1$, se cumple que*

$$f(\lambda\overline{x} + (1-\lambda)\overline{y}) \geq \lambda f(\overline{x}) + (1-\lambda)f(\overline{y}).$$

Definición* 1.13.** *Se dice que f es* ***estrictamente cóncava *si se cumple*

$$f(\lambda\overline{x} + (1-\lambda)\overline{y}) > \lambda f(\overline{x}) + (1-\lambda)f(\overline{y})$$

para todo $\overline{x}$, $\overline{y} \in D$ con $\overline{x} \neq \overline{y}$ y $0 < \lambda < 1$.

***Definición* 1.14.** *Un conjunto $C \subset \mathbb{R}^n$ no vacío es convexo si, para cualquiera dos puntos distintos x e y en C, y para cualquier número real α en el intervalo $(0, 1)$, el punto $(1-\alpha)x + \alpha y$ también está en C, es decir, C está cerrado bajo combinaciones convexas de dos miembros cualesquiera de C.*

El siguiente resultado relaciona la concavidad y convexidad de la función con sus valores máximos y mínimos.

***Teorema* 1.4.** *Sean $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ una función de clase C^2 definida en el conjunto convexo y abierto D y sea $\bar{x}^* \in D$ un punto crítico. Entonces:*

1. *Si f es una función convexa en D, $\bar{x}^*$ es un mínimo global de f. Además, si la función es estrictamente convexa el mínimo es estricto.*

2. *Si f es una función cóncava en D, $\bar{x}^*$ es un máximo global de f. Además, si la función es estrictamente cóncava el máximo es estricto.*

Para su demostración ver [16].

Los siguientes teoremas serán utilizados dentro en los capítulos 3 y 4.

Teorema* 1.5. *(Teorema Weierstrass)
Sea $D \subset \mathbb{R}$ compacto y $f : D \longrightarrow \mathbb{R}$ una función continua en D. Entonces f alcanza un máximo y un mínimo en D, es decir, existen z_1 y z_2 puntos en D tal que

$$f(z_2) \leq f(x) \leq f(z_1),\ x \in D.$$

Para su demostración ver [15].

Teorema* 1.6. *(Teorema de la Función implícita)
Sea $F : S \subset \mathbb{R}^{m+n} \longrightarrow \mathbb{R}^n$, donde H es una función de clase C^1 y S abierto , sea (x^*, y^*) un punto en S tal que $D_yF(x^*, y^*)$ es invertible, y sea $F(x^*, y^*) = c$. Entonces, hay una vecindad $N \subset \mathbb{R}^m$ de x^*, y una función $\xi : N \longrightarrow \mathbb{R}^n$ de clase C^1 que cumplen:

1. $((x, \xi(x)) \in S \quad \forall x \in N$.
2. $\xi(x^*) = y^*$.
3. $F(x, \xi(x)) = c \quad \forall x \in N$.

Para su demostración ver [15].

***Teorema* 1.7.** *Sea $f : \mathbb{R}^l \longrightarrow \mathbb{R}$ es una función continua, si $f(x)$ tiene derivada en el conjunto convexo $D \subset \mathbb{R}^l$, entonces*

1. *La función $f(x)$ es convexa si y sólo si*

$$f(x) + \bigtriangledown f(x)(y - x) \leq f(y)$$

para todo $x \in D$ y $y \in D$.

2. *La función $f(x)$ es estrictamente convexa en D si y sólo si*

$$f(x) + \bigtriangledown f(x)(y - x) < f(y).$$

para todo $x \in D$ y $y \in D$ con $x \neq y$.

Para su demostración ver [13].

***Teorema* 1.8.** *Sea $g : C \subset \mathbb{R}^l \longrightarrow \mathbb{R}$ una función convexa, si x_0 es un punto interior del conjunto convexo C, entonces hay un vector u distinto de cero, tal que*

$$g(x) \geq g(x_0) + \langle u, x - x_0 \rangle.$$

Para su demostración ver [5].

En la siguiente imagen podemos observar la representación gráfica de la función $g(x)$, para x_0 y $x \in C$.

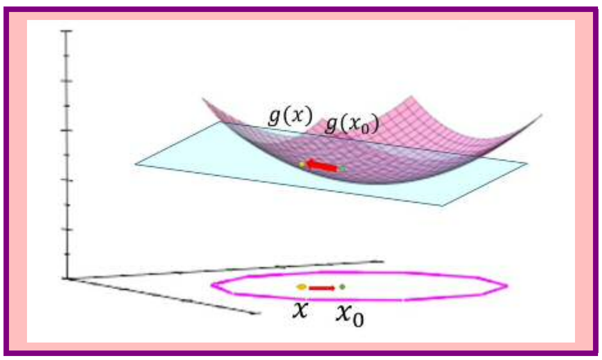

Figura 1.2: Representación gráfica del Teorema 1.8.

Capítulo 2

Método de los Multiplicadores de Lagrange.

2.1. Joseph Louis Lagrange.

El interés de Lagrange por las matemáticas comenzó cuando leyó una copia de la obra de Halley de 1693 sobre el uso del álgebra en la óptica. También se sintió atraído por la física en el Colegio de Turín y decidió hacer una carrera en matemáticas. El 23 de julio de 1754 publicó su primer trabajo matemático, el artículo hace una analogía entre el teorema binomial y las derivadas sucesivas del producto de funciones.

Comenzó a trabajar en la tautocrona, la curva en la que una partícula ponderada siempre llegará a un punto fijo al mismo tiempo, independientemente de su posición inicial. A fines de 1754 había hecho algunos descubrimientos importantes sobre la tautocrona que contribuirían sustancialmente al nuevo tema del cálculo de variaciones. Lagrange envió a Euler sus resultados sobre la tautocrona que contiene su método de máximos y mínimos.

Los artículos de Lagrange cubren una variedad de temas. Publicó sus resultados sobre el cálculo de variaciones y un breve trabajo sobre el cálculo de probabilidades. En un trabajo sobre los fundamentos de la dinámica, Lagrange basó su desarrollo en el principio de la acción mínima y en la energía cinética. En Mélanges de Turin, Lagrange también realizó un estudio importante sobre la propagación del sonido, e hizo importantes contribuciones a la teoría de las cuerdas vibrantes.

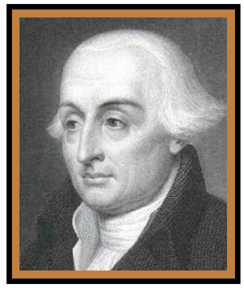

Figura 2.1: Joseph-Louis Lagrange (1736 - 1813)

2.2. Función de Lagrange.

Un problema de optimización con restricciones de igualdad es maximizar una función $f(\overline{x})$ sujeto a $\overline{x} \in D = U \cap \{x \mid g(x) = 0\}$, donde $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ y $g : \mathbb{R}^n \longrightarrow \mathbb{R}^k$ son funciones de clase C^1 con $g(\bar{x}) = (g_1(\bar{x}), ..., g_k(\bar{x}))$ y $U \subset \mathbb{R}^n$ abierto. A la condición sobre el dominio D se le llama restricción de igualdad restringida debido a que $g(x) = 0$.

Definamos $L : D \times \mathbb{R}^k \longrightarrow \mathbb{R}$, **la función de Lagrange** como:

$$L(\bar{x}, \lambda) = f(\bar{x}) + \sum_{i=1}^{k} \lambda_i g_i(\bar{x}).$$

El vector $\lambda = (\lambda_1, ..., \lambda_k) \in \mathbb{R}^k$ se le conoce como el vector de **los multiplicadores de Lagrange**.

Para mayor información ver [2], [14] y [11].

2.3. Demostración del Teorema de Lagrange.

En adelante sólo denotaremos a x como el vector $\bar{x}$ en su $\mathbb{R}^n$ respectivo.

Teorema de Lagrange. *Sean $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ y $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, funciones de clase C^1 para $i = 1, \ldots, k$. Supongamos que x^* es un máximo o mínimo de f en el conjunto*

$$D = U \cap \left\{ x \mid g_i(x) = 0, i = 1, \ldots, k \right\},$$

donde $U \subset \mathbb{R}^n$ abierto y $g(x) = (g_1(x), ..., g_k(x))$. Supongamos también $\rho(Dg(x^*)) = k$. Entonces existe un vector $\lambda^* = (\lambda_1^*, \ldots, \lambda_k^*) \in \mathbb{R}^k$ de tal manera que

$$Df(x^*) + \sum_{i=1}^{k} \lambda_i^* Dg_i(x^*) = 0.$$

El valor $\rho(Dg(x^*)) = k$ denota el rango de la matriz $Dg(x^*)$.

Demostración.

La siguiente simplificación de notación ayudará en gran medida a la prueba del teorema.

Supongamos sin pérdida de generalidad, que la submatriz $k \times k$ de $Dg(x^*)$ de rango completo es la submatriz $k \times k$ que consiste en las primeras k filas y k columnas.

Denotaremos las primeras k coordenadas de un vector $x \in D$ por w, y las últimas $n-k$ coordenadas por z, es decir, escribimos $x = (w, z)$. En particular, escribiremos $x^* = (w^*, z^*)$ para denotar el óptimo local x^*.

También denotaremos por $D_w f(w, z)$, a la derivada de f en (w, z) con respecto a la variable w únicamente, y por $D_z f(w, z)$, a la derivada de f en (w, z) con respecto a la variable z. $D_w g(w, z)$ y $D_z g(w, z)$ se definen de manera similar. Tomemos en cuenta que las dimensiones de las matrices $D_w f(w, z)$, $D_z f(w, z)$, $D_w g(w, z)$ y $D_z g(w, z)$ son respectivamente $1 \times k$, $1 \times (n - k)$, $k \times k$ y $k \times (n - k)$.

Trataremos al vector λ^* en $\mathbb{R}^k$, como una matriz de $1 \times k$. Así, por ejemplo, escribiremos $\lambda^* Dg(x^*)$ para representar $\sum_{i=1}^{k} \lambda_i^* Dg_i(x^*)$.

Con esta notación $x = (w^*, z^*)$ es un máximo local de f en $D = U \cap \left\{ (w, z) \in \mathbb{R}^n \mid g(w, z) = 0 \right\}$ y $\rho(D_w g(w^*, z^*)) = k$. Ahora debemos demostrar que existe λ^* tal que:

$$D_w f(w^*, z^*) + \lambda^* D_w g(w^*, z^*) = 0,$$
$$D_z f(w^*, z^*) + \lambda^* D_z g(w^*, z^*) = 0.$$

Dado que $\rho(D_w g(w^*, z^*)) = k$, el teorema de la función implícita muestra que existe un conjunto abierto V en $\mathbb{R}^{n-k}$ que contiene a z^*, y una función $h : V \longrightarrow \mathbb{R}^k$ de clase C^1 tal que $h(z^*) = w^*$ y $g(h(z), z) \equiv 0$ para todo $z \in V$. Diferenciando la identidad $g(h(z), z) \equiv 0$ con respecto a z utilizando la regla de la cadena, obtenemos.

$$D_w g(h(z), z) Dh_z + D_z g(h(z), z) = 0$$

Por hipótesis sobre el rango tenemos que $D_w g(w^*, z^*)$ es invertible, esto implica que

$$Dh(z^*) = -\left[D_w g(w^*, z^*)\right]^{-1} D_z g(w^*, z^*).$$

Ahora definimos λ^* por

$$\lambda^* = -D_w f(w^*, z^*) \left[D_w g(w^*, z^*)\right]^{-1}.$$

Mostraremos que λ^* así definido cumple las condiciones requeridas. De hecho, de la misma definición de λ^* multiplicaremos ambos lados por $D_w g(w^*, z^*)$ y obtenemos,

$$\begin{aligned}\lambda^* D_w g(w^*, z^*) &= -D_w f(w^*, z^*) \left[D_w g(w^*, z^*)\right]^{-1} D_w g(w^*, z^*) \\ &= -D_w f(w^*, z^*).\end{aligned}$$

Es decir

$$D_w f(w^*, z^*) + \lambda^* D_w g(w^*, z^*) = 0.$$

Aún queda por demostrar que

$$D_z f(w^*, z^*) + \lambda^* D_z g(w^*, z^*) = 0.$$

Definamos la función $F : V \longrightarrow \mathbb{R}$ por $F(z) = f(h(z), z)$. Como f tiene un mínimo o máximo local en $(w^*, z^*) = (h(z^*), z^*)$, es inmediato que F tiene un máximo local en z^*. Dado que V es abierto, z^* es un máximo local sin restricciones de F, y $DF(z^*) = 0$, o bien

$$D_w f(w^*, z^*) Dh(z^*) + D_z f(w^*, z^*) = 0.$$

Sustituyendo a $Dh(z^*)$, obtenemos

$$-D_w f(w^*, z^*) \left[D_w g(w^*, z^*)\right]^{-1} D_z g(w^*, z^*) + D_z f(w^*, z^*) = 0,$$

que a partir de la definición de λ^*, es igual a

$$D_z f(w^*, z^*) + \lambda^* D_z g(w^*, z^*) = 0.$$

■

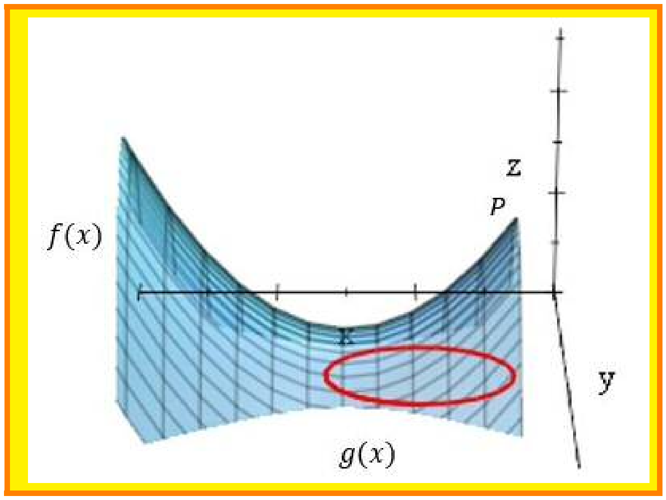

Figura 2.2: Representación gráfica del Teorema de Lagrange, con la función objetivo $f(x)$ sujeta a la restricción $g(x)$.

La Figura 2.2 muestra a la función $f(x)$ y $g(x)$, donde P es el máximo de la función $f(x)$ restringido a $g(x) = 0$.

En la siguiente seccíon se mostrarán algunos ejemplos que ilustran la aplicación del teorema anterior, también conocido como Método de Lagrange.

2.4. Aplicaciones del Método de Lagrange.

Para ilustrarar el Teorema de Lagrange mostratemos los siguientes ejemplos.

Ejemplo. *Sea* $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *dada por,* $f(x, y) = 2x + 3y$*, restringida a* $x^2 + y^2 = 4$*.*

Para poder encontrar donde se alcanza el máximo y mínimo de la función sujeta a la restricción, recordemos primero que la función de Lagrange está definida, por:

$$L(x, y, \lambda) = f(x, y) + \sum_{i=1}^{k} \lambda_i g_i(x, y),$$

donde tenemos como única restricción ($k = 1$) a la función $g(x, y) = x^2 + y^2 - 4$, así,

$$L(x, y, \lambda) = (2x + 3y) + \lambda \left(x^2 + y^2 - 4\right).$$

Utilizando el teorema de Lagrange se tiene que existe $\lambda \geq 0$ tal que

$$Df(x, y) + \sum_{i=1}^{k} \lambda_i Dg_i(x) = (2, 3) + \lambda(2x, 2y) = 0.$$

Los puntos críticos son las soluciones del sistema

$$\begin{aligned} 2 + 2\lambda x &= 0, & (1)\\ 3 + 2\lambda y &= 0, & (2)\\ x^2 + y^2 - 4 &= 0. & (3) \end{aligned}$$

Resolviendo el sistema de ecuaciones encontramos el valor de las variables x y y, donde tenemos el punto crítico $(x, y) = \left(\pm\frac{4}{\sqrt{13}}, \pm\frac{6}{\sqrt{13}}\right)$. Evaluando en nuestra función

$$f\left(\frac{4}{\sqrt{13}}, \frac{6}{\sqrt{13}}\right) = \frac{8}{\sqrt{13}} + \frac{18}{\sqrt{13}} = \frac{26}{\sqrt{13}},$$

$$f\left(-\frac{4}{\sqrt{13}}, -\frac{6}{\sqrt{13}}\right) = -\frac{8}{\sqrt{13}} - \frac{18}{\sqrt{13}} = -\frac{26}{\sqrt{13}}.$$

Por lo tanto el valor máximo $\frac{26}{\sqrt{13}}$ se alcanza en $\left(\frac{4}{\sqrt{13}}, \frac{6}{\sqrt{13}}\right)$ mientras que el valor mínimo es $-\frac{26}{\sqrt{13}}$ y se alcanza en $\left(-\frac{4}{\sqrt{13}}, -\frac{6}{\sqrt{13}}\right)$.

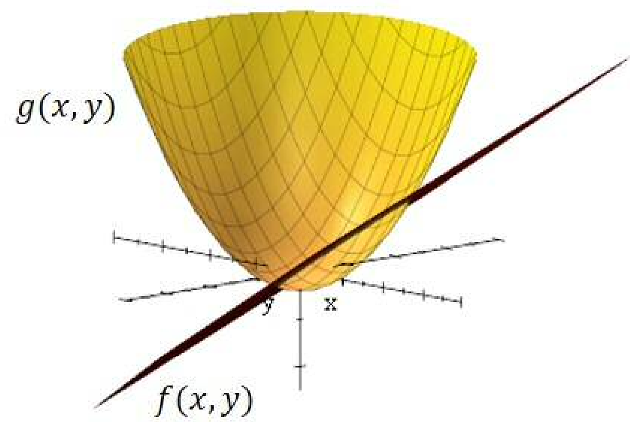

Figura 2.3: Representación de $f(x,y)$ restringida a $g(x,y)=0$

En la Figura 2.3 podemos apreciar donde se alcanzan el máximo y mínimo de la función $f(x,y)$ restringida a $g(x,y)$.

Ejemplo. *Aplicación a la economía. Ahora consideremos el caso de una empresa que fabrica dos artículos en cantidades x, y, donde la función de costos está dada por $C(x,y) = x^2 + 2y^2 + xy + 20$ y el presupuesto es \$ 8770.*

Para saber cual es la producción máxima que puede alcanzar dicha empresa, definamos $f(x,y) = x + y$ y $g(x,y) = 8770 - (x^2 + 2y^2 + xy + 20)$. Así la función de Lagrange es:

$$\begin{aligned} L(x,y,\lambda) &= (x+y) + \lambda\left(8770 - \left(x^2 + 2y^2 + xy + 20\right)\right) \\ &= (x+y) + \lambda\left(8750 - x^2 - 2y^2 - xy\right). \end{aligned}$$

Utilizando el Teorema de Lagrange

$$(1,1) + \lambda(-2x - y, -4y - x) = 0.$$

Los puntos críticos son las soluciones del sistema

$$\begin{aligned} 1 - 2x\lambda - y\lambda &= 0, \qquad (1) \\ 1 - 4y\lambda - x\lambda &= 0, \qquad (2) \\ 8750 - x^2 - 2y^2 - xy &= 0. \qquad (3) \end{aligned}$$

Despejando a λ de la ecuación (1) y (2), se tiene:

$$\lambda = \frac{1}{2x+y},\ \lambda = \frac{1}{4y+x},$$

e igualando las expresiones resultantes, obtenemos $x = 3y$. Sustituyendo en (3) se tiene $14y^2 = 8750$, con lo que $y = 25$ o $y = -25$. Rechazamos este último resultado por carecer de interpretación económica y sustituimos en los resultados anteriores, con lo que al final obtenemos un punto crítico para este problema que es $(75, 25)$ con multiplicador asociado de $\lambda = \dfrac{1}{175}$. Como la restricción es una ecuación de grado 2 en dos variables entonces es aplicable el Teorema 1.3, para esto la matriz Hessiana respecto de las variables x, y está dada por:

$$H_{(x,y)}f(x,y) = \begin{pmatrix} -2\lambda & -\lambda \\ -\lambda & -4\lambda \end{pmatrix},$$

que calculada en el punto crítico $(75, 25)$ asociado al multiplicador $\dfrac{1}{175}$ es

$$H_{(x,y)}f(75,25) = \begin{pmatrix} -\frac{2}{175} & -\frac{1}{175} \\ \\ -\frac{1}{175} & -\frac{4}{175} \end{pmatrix}.$$

Así los menores conducentes son:

$$A_1 = -\frac{2}{175} < 0, \quad A_1 = \frac{1}{4375} > 0.$$

Por el Teorema 1.1 y Teorema 1.3 existe un máximo local relativo en el punto $(75, 25)$ con multiplicador asociado $\lambda = \dfrac{1}{175}$.

Así el máximo nivel de producción que puede alcanzar la empresa es $Q(75, 25) = 100$ unidades de producto.

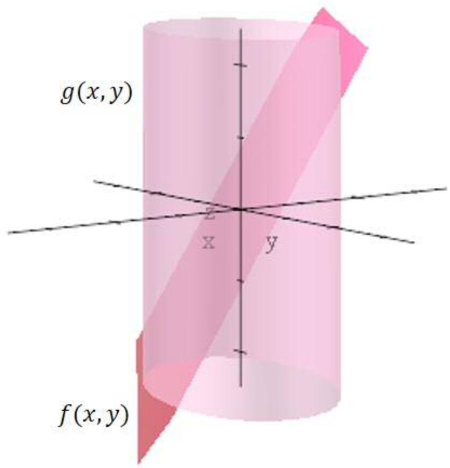

Figura 2.4: Representación de la función $f(x,y)$ restringida a $g(x,y)=0$

En la Figura 2.4 se muestra donde la función $f(x,y)$ alcanza su máximo, sujeta a la restricción $g(x,y)=0$.

Ejemplo. *Una empresa elaborará un producto, con* $f(L,K)=50L^{\frac{2}{3}}K^{\frac{1}{3}}$ *donde* L *representa las unidades de mano de obra y* K *las unidades de capital. Le cuesta a la empresa* \$100 *por cada unidad de mano de obra y* \$300 *por cada unidad de capital empleado. La empresa dispone de* \$45000 *para propósitos de producción.*

Para saber donde se da la máxima producción sujeta a nuestra restricción, damos en este caso $g(L,K)=45000-100L-300K$.
La función de Lagrange es

$$L(L,K,\lambda)=50L^{\frac{2}{3}}K^{\frac{1}{3}}+\lambda\,(45000-100L-300K)$$

mientras que por Teorema de Lagrange tenemos

$$\left(\frac{100}{3}L^{\frac{-1}{3}}K^{\frac{1}{3}},\frac{50}{3}L^{\frac{2}{3}}K^{\frac{-2}{3}}\right)+\lambda\,(-100,-300) \;=\; 0.$$

Así tenemos el sistema de ecuaciones

$$\begin{aligned}
\frac{100}{3}L^{\frac{-1}{3}}K^{\frac{1}{3}}-100\lambda &= 0, && (1)\\
\frac{50}{3}L^{\frac{2}{3}}K^{\frac{-2}{3}}-300\lambda &= 0, && (2)\\
45000-100L-300K &= 0. && (3)
\end{aligned}$$

Despejando a λ de las ecuaciones (1) y (2), se tiene

$$\lambda = \frac{1}{3}L^{\frac{-1}{3}}K^{\frac{1}{3}} \quad \text{y} \quad \lambda = \frac{1}{18}L^{\frac{2}{3}}K^{\frac{-2}{3}}.$$

Igualando ambas expresiones y despejando a L,

$$L = 6K.$$

Sustituyendo L en la ecuación (3) tenemos

$$45000 - 100(6K) - 300K = 0.$$

De donde se tiene $K = 50$ y sustituyendo en L, tenemos $L = 300$. La empresa maximiza su producción si emplea 300 unidades de mano de obra y 50 de capital.

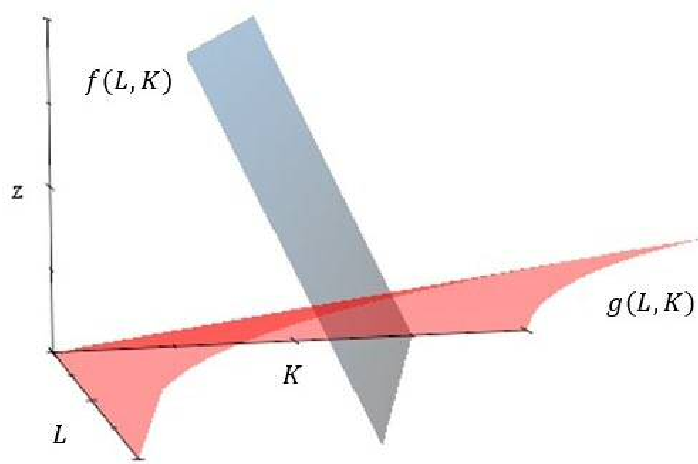

Figura 2.5: Representación de la función $f(L,K)$ sujeto a la restricción $g(L,K) = 0$

En la Figura 2.5 se muestra donde la función $f(L,K)$ alcanza su máximo sujeto a la restricción $g(L,K)$.

Para mayor información ver [2], [8], [9] y [12].

Capítulo 3

Teorema de Kuhn y Tucker.

3.1. Albert W. Tucker y Harold William Kuhn.

Albert W. Tucker matemático canadiense, nacido el 28 de noviembre, 1905 y muerto el 25 de enero de 1995. Graduado de la Universidad de Toronto en 1929, desarrolló su carrera en los Estados Unidos. Sentó las bases de la programación lineal y desarrolló la teoría de juegos, siendo inventor del dilema de los prisioneros, creado como una ayuda educativa para los estudiantes de psicología de Stanford.

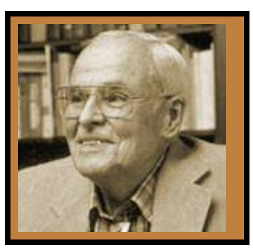

Figura 3.1: Albert W. Tucker

Harold William Kuhn nacido en Santa Mónica, California, 29 de julio de 1925, fue un matemático estadounidense que estudió teoría de juegos, ganó el Premio de Teoría John von Neumann en 1980 junto con David Gale y Albert W. Tucker. Profesor emérito de matemáticas en la Universidad de

Princeton. Fue conocido por su asociación con John Forbes Nash, un amigo de toda la vida y colega, y una figura clave para lograr que Nash ganara la atención del comité que lo llevó a obtener el premio novel en economía en 1994.

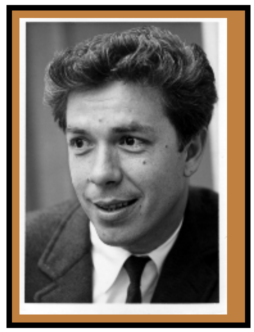

Figura 3.2: Harold William Kuhn

Las condiciones de Kuhn y Tucker fueron desarrolladas por Albert William Tucker y complementada por Harold Kuhn, quien permitió mejoras en el proceso, pero se le adjudicó un papel secundario al Teorema de Kuhn y Tucker que busca generalizar el método de Multiplicadores de Lagrange, este método consiste en resolver el problema no lineal como uno sin restricciones utilizando la función de Lagrange. Una vez hecho esto, si la solución óptima encontrada no cumple la totalidad o parte de las restricciones del problema se aplican tales restricciones, y se resuelve nuevamente. Esto se repite hasta llegar a un conjunto de restricciones activas cuya solución satisface también las restricciones omitidas.

3.2. Teorema de Kuhn y Tucker.

Consideremos el conjunto de restricciones de la forma:

$$D = U \cap \{x \in \mathbb{R}^n | h_i(x) \geqslant 0, \quad i = 1, \ldots, l\},$$

donde $U \subset \mathbb{R}^n$ abierto, y $h_i : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad i = 1, \ldots, l$. El Teorema de Lagrange se puede combinar con la teoría de Kuhn y Tucker para obtener las condiciones óptimas locales en el caso general de problemas de optimización definida por las restricciones mixtas, donde el conjunto de restricciones está dada por:

$$D = U \cap \{x \in \mathbb{R}^n | g_i(x) = 0, \quad j = 1, \ldots, k, \quad h_i(x) \geqslant 0, \quad i = 1, \ldots, l\}.$$

El Teorema de Kuhn y Tucker proporciona una elegante caracterización del comportamiento de la función objetivo f y las funciones de restricción h_i en el óptimo local de problemas de optimización restringidos por la desigualdad. Las condiciones que describe se pueden ver como las condiciones necesarias de primer orden para obtener el óptimo local en estos problemas.

Diremos que una restricción de desigualdad $h_i(x) \geqslant 0$ es **efectiva** para el punto x^* si $h_i(x^*) = 0$.

Utilizamos la expresión $|E|$ para denotar la cardinalidad de un conjunto finito E, es decir, el número de elementos en el conjunto E.

Teorema de Kuhn y Tucker. *Sean $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ y $h_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, funciones de clase C^1 para $i = 1, \ldots, l$. Supongamos que x^* es un máximo local de f en el conjunto*

$$D = U \cap \left\{ x \in \mathbb{R}^n \mid h_i(x) \geq 0, i = 1, \ldots, l \right\},$$

donde U es un conjunto abierto en $\mathbb{R}^n$. Si $E \subset \{1, \ldots, l\}$ denota el conjunto de restricciones en x^* y si $h_E = (h_i)_{i \in E}$. Supongamos que $\rho(Dh_E(x^*)) = |E|$, entonces existe un vector $\lambda^* = (\lambda_1^*, \ldots, \lambda_l^*) \in \mathbb{R}^l$ de tal modo que se cumplen las siguientes condiciones:

1. $\lambda_i^* \geq 0 \quad y \quad \lambda_i^* h_i(x^*) = 0 \qquad \forall i = 1 \ldots l,$

2. $Df(x^*) + \sum_{i=1}^{l} \lambda_i^* Dh_i(x^*) = 0.$

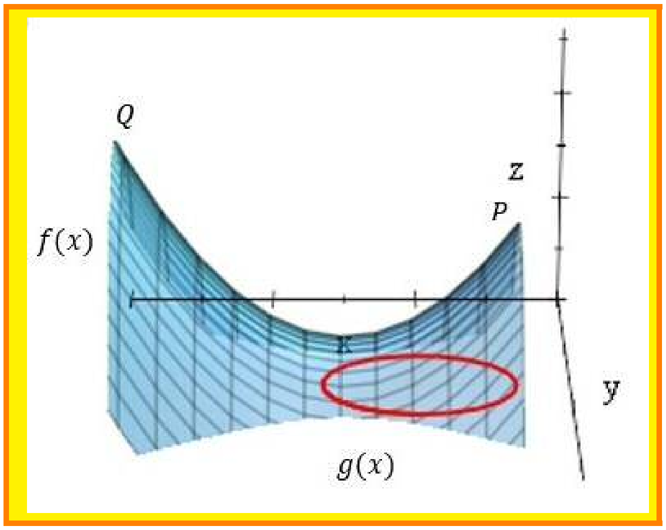

Figura 3.3: Representación Gráfica del Teorema de Kuhn y Tucker, con la función objetivo $f(x)$ y una restricción $g(x)$

La Figura 3.3 muestra a la función $f(x)$ y $g(x)$, con $g(x) = h(x)$ ya que en este caso se encuentra una sola restricción, Q es el máximo de la función $f(x)$ restringido a la región definida por la desigualdad $g(x) \geq 0$, note que si P es el máximo de la función $f(x)$ restringido a la igualdad $g(x) = 0$ dado por el Teorema de Lagrange entonces, $P \neq Q$.

Aunque hemos establecido el Teorema de Kuhn y Tucker para los máximos locales el teorema se puede extender fácilmente para cubrir los mínimos locales. Porque si x^* fuera un mínimo local de f en D, x^* sería un máximo local de $-f$ en D y dado que $D(-f) = -Df$ tendríamos lo siguiente:

***Corolario* 3.1.** *Supongamos que f y D se definen como en el teorema de Kuhn y Tucker y x^* es un mínimo local de f en D. Sea E el conjunto de restricciones efectivas en x^*, si $h_E = (h_i)_{i \in E}$ y supóngase que $\rho(Dh_E(x^*)) = |E|$, entonces existe $\lambda^* \in \mathbb{R}^l$ tal que*

1. $\lambda_i^* \geq 0 \quad y \quad \lambda_i^* h_i(x^*) = 0 \qquad \forall i,$
2. $Df(x^*) - \sum_{i=1}^{l} \lambda_i^* Dh_i(x^*) = 0.$

Diremos que el par (x^*, λ^*) cumplen las **condiciones necesarias de primer orden para un máximo** (o cumplen las condiciones de primer orden de Kuhn y Tucker para un máximo), si (x^*, λ^*) satisface $\lambda^* \geqslant 0$ así como las condiciones del teorema anterior. Del mismo modo diremos que (x^*, λ^*) cumplen las **condiciones necesarias de primer orden para un mínimo** (o que cumplen las condiciones necesarias de primer orden de Kuhn y Tucker para un mínimo) si satisface $\lambda^* \geqslant 0$ así como las condiciones anteriores del corolario.

3.3. Primera condición del teorema.

La condición $\lambda_i^* \geq 0 \quad \text{y} \quad \lambda_i^* h_i(x^*) = 0 \quad \forall i = 1 \ldots l$, es llamada condición de “holgura complementaria". La terminología surge de la observación de que debemos tener $\lambda_i^* = 0$ si $h_i(x)^* > 0$, y $h_i(x^*) = 0$ si $\lambda_i^* > 0$, es decir, la desigualdad no es estricta.

Debemos tener en cuenta que el Teorema de Kuhn y Tucker proporciona condiciones que son solo esenciales para los óptimos locales. Estas condiciones se requieren para determinar un punto como óptimo local y de hecho es fácil incluir ejemplos para demostrar que las condiciones pueden no ser suficientes.

Veamos uno particularmente simple.

Ejemplo. *Sea $f : \mathbb{R} \to \mathbb{R}$ y $g : \mathbb{R} \to \mathbb{R}$ definidas por $f(x) = x^3$ y $g(x) = x$ respectivamente. Consideremos el problema de maximizar f en el conjunto $D = \{x \in \mathbb{R} | g(x) \geqslant 0\}$*

Sea $x^* = \lambda^* = 0$. Entonces la única restricción del problema es efectiva en x^*. De la restriccion de que $g(x) = x$ entonces el rango es 1, supongamos que se cumple $f^{'}(x^*) + \lambda^* g^{'}(x^*) = 0$, si las condiciones del Teorema de Kuhn y Tucker también fueran suficientes, x^* sería un máximo local de f en D. Sin embargo, f está aumentando estrictamente en $x^* = 0$. Por tanto x^* no puede ser un máximo local.

3.4. La cualificación de restricción.

Al igual que la condición análoga del Teorema de Lagrange la conclusión del Teorema de Kuhn y Tucher de que el rango de $Dh_E(x^*)$ es igual a $|E|$ se denomina **cualificación de restricción**. Esta condición juega un papel central en la prueba del teorema. Además si falla la cualificación de la restricción, el teorema mismo podría fallar. Aquí hay un ejemplo para ilustrar este punto.

Ejemplo. *Sea* $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *y* $h : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *definida por* $f(x,y) = -(x^2+y^2)$ *y* $h(x,y) = (x-1)^3 - y^2$. *Consideremos el problema de maximizar* f *en el conjunto*

$$D = \left\{(x,y) \in \mathbb{R}^2 | h(x,y) \geq 0\right\}.$$

Una solución a este problema se puede obtener mediante inspección, la función f alcanza un máximo en el punto donde $x^2 + y^2$ alcanza un mínimo. Dado que la restricción requiere

$$(x-1)^3 \geq y^2, \text{ y } y^2 \geq 0 \ \forall y,$$

el valor mas pequeño de x en D es $x = 1$ lo que ocurre cuando $y = 0$. Se deduce que f se maximiza en $(x^*, y^*) = (1, 0)$. Tengamos en cuenta que la restricción es única.

En este máximo global de f en D, tenemos $Dh(x^*, y^*) = (3(x^*-1)^2, 2y^*)) = (0, 0)$ entonces $\rho(Dh(x^*, y^*)) = 0 < 1$, y la cualificación de restricción falla. Por otra lado tenemos $Df(x^*, y^*) = (-2x^*, -2y^*) = (-2, 0)$ por lo que no puede existir $\lambda \geq 0$ tal que $Df(x^*, y^*) + \lambda Dh(x^*, y^*) = (0, 0)$. Con esto útimo observamos que las conclusiones del Teorema de Kuhn y Tucker también fallan.

Multiplicadores de Kuhn y Tucker

El vector λ^* en el Teorema de Kuhn y Tucker se llama **vector de multiplicadores** de Kuhn y Tucker correspondiente al máximo local x^*. Al igual que con los multiplicadores de Lagrange, también se puede pensar que los multiplicadores de Kuhn y Tucker miden la sensibilidad de la función objetivo en x^*.

De hecho esta interpretación es particularmente intuitiva en el contexto de las restricciones de primer orden. Es decir, si $h_i(x^*) > 0$, entonces la restricción i-ésima restringirá a elevar el valor de la función objetivo en el ejercicio de maximización y λ^* deberá ser cero. Por otro lado si $h_i(x^*) = 0$ entonces debilitar la restricción i-ésima puede ayudar a aumentar el valor del ejercicio de maximización y entonces tendremos $\lambda_i \geq 0$.

Para visualizar esta interpretación de λ, supongamos que las funciones de restricción h_i se dan todas de forma paramétrica como

$$h_i(x; c) = h_i(x) + c_i.$$

Bajo esta suposición la i-ésima restricción cambia conforme al valor del parámetro c_i. Sea $C \subset \mathbb{R}^l$ es un conjunto abierto de valores factibles para los parámetros $c = (c_1, \ldots, c_l)$. Supongamos que para cada $c \in C$, hay un máximo global $x^*(c)$ de f en el conjunto de restricciones

$$D(c) = U \cap \left\{ x \in \mathbb{R}^n | h_i(x) + c_i \geq 0, i = 1, \ldots, l \right\}.$$

Supongamos además que la restricción de cualificación se mantiene en cada $c \in C$ entonces, de manera análoga como se hizo en el Teorema de Lagrange existe $\lambda^*(c) \geq 0$ tal que

$$Df(x^*(c)) + \sum_{i=1}^{l} \lambda_i^* Dh_i(x^*(c)) = 0.$$

Finalmente supongamos que λ^* y x^* son funciones de clase C^1 con parámetros sobre c en el conjunto C.

Definamos $F : C \longrightarrow \mathbb{R}$ por

$$F(c) = f(x^*(c)).$$

si afirmamos que $\partial F(c)/\partial c_i = \lambda_i$ para $i = 1, \ldots, l$. Eligiendo cualquier $c \in C$, supongamos i tal que $h_i(x) + c_i > 0$, sea $\tilde{c}_i$ tal que $\tilde{c}_i < c_i$ y

$$h_i(x) + \tilde{c}_i > 0.$$

Considerando el conjunto de restricciones $D(\tilde{c}_i)$ que resulta cuando el parámetro c_i en la restricción i se remplaza por $\tilde{c}_i < c_i$, debemos obtener

$$D(\tilde{c}_i) \subset D(c).$$

Como $x^* = x^*(c)$ es un máximo local de f en el conjunto de restricciones más grande $D(c)$ y como $x^*(c) \in D(c_i)$ entonces $h_i(x^*(c) + c_i) \geq 0 \; \forall \, i$. Por elección de $\tilde{c}_i$ se tiene $h_i(x^*(c) + \tilde{c}_i) \geq 0$ y $h_i(x^*(c)) + c_i \geq 0$ por tanto $x^*(c) \in D(\tilde{c}_i)$, con esto se tiene que $x^*(c)$ también es un máximo de f en el conjunto de restricciones $D(\tilde{c}_i)$. Por tanto

$$F(\tilde{c}_i) = f(x^*(c)) = F(c).$$

Como c se eligió de manera arbitraria, podemos afirmar que F es constante, de donde se deduce que $\partial F(c)/\partial c_i = 0$. Por otro lado también se cumple que $h_i(x^*(c)) + c_i > 0$, lo cual implica que se cumplen las condiciones de Kuhn y Tucker de "holguras complementarias" $\lambda_i^*(c) = 0$. Por lo tanto

$$\frac{\partial F}{\partial c_i}(c) = \lambda_i^*(c).$$

Para mayor información ver [3], [4], [5], [6], [10] y [14].

3.5. Aplicación del teorema.

3.5.1. Procedimiento.

La forma de aplicar el Teorema de Kuhn y Tucker para resolver un problema de optimización restringido por la desigualdad, implica esencialmente los mismos pasos que el uso del Teorema de Lagrange para resolver problemas con restricciones de igualdad. Es decir, consideremos una función Lagrangiana luego calculamos sus puntos críticos y finalmente, evaluamos la función objeto en cada punto crítico y seleccionamos el punto en el que se optimiza la función.

Existen diferencias muy importantes en los detalles. Uno de ellos surge del hecho de que las conclusiones del Teorema de Kuhn y Tucker son direntes para los máximos locales y mínimos locales. Esto requiere una diferencia en los pasos a seguir en la resolución de problemas de maximización para poder resolver problemas de minimización.

Proceso de maximización de desigualdad restringida.

$$\text{Maximizar } f(x) \text{ sujeto a } x \in D = U \cap \{x | h(x) \geq 0\}.$$

Paso 1.
Como primer paso en el procedimiento para resolver este problema definiremos una función $L : \mathbb{R}^n \times \mathbb{R}^l \longrightarrow \mathbb{R}$ que seguiremos llamando Lagrangiana,

dada por

$$L(x,\lambda) = f(x) + \sum_{i=1}^{l} \lambda_i h_i(x).$$

Paso 2.

El segundo paso en el procedimiento es encontrar toda las soluciones (x,λ) al siguiente conjunto de vectores

$$\frac{\partial L}{\partial x_j}(x,\lambda) = 0, \qquad j = 1,\ldots,n.$$

$$\frac{\partial L}{\partial \lambda_i}(x,\lambda) \geq 0, \qquad \lambda_i \geq 0, \qquad \lambda_i \frac{\partial L}{\partial \lambda_i}(x,\lambda) = 0, \qquad i = 1,\ldots,l.$$

Cualquier solución al sistema de ecuaciones se llamara “punto crítico"de L. Es importante tener en cuenta que las condiciones que definen los puntos críticos de L difieren a las correspondientes en los problemas de igualdad restringida, en particular con respecto a las derivadas respecto a λ. Vamos a denotar por M al conjunto de puntos críticos de L, para los que $x \in U$.

$$M = \{(x,\lambda) | (x,\lambda) \text{ es un punto crítico de L y } x \in U\}.$$

Paso 3.

Como tercer y último paso, calculemos el valor de f en cada $x \in M$. En la práctica el valor de x que maximiza f sobre este conjunto suele ser también la solución del problema de maximización original.

Proceso de minimización de desigualdad restringida.

Supongamos que el problema original es resolver:

$$\text{minimizar } f(x) \text{ sujeto a } x \in D = U \cap \{x | x,\, h(x) \geq 0\}.$$

Dos rutas están abiertas para nosotros:

Dado que x maximiza f sobre D si y sólo si x minimiza $-f$ sobre D, podríamos simplemente reescribir el problema como un problema de minimización con la función objetivo dada por $-f$ y usar el procedimiento anterior.

Las rutas alternativas requieren modificar la función de Lagrange de tal modo que

$$L(x,\lambda) = f(x) - \sum_{i=1}^{l} \lambda_i h_i(x),$$

y seguimos los mismos pasos restantes que se enumeran para el problema de maximización, es decir, en el segundo paso encontraremos el conjunto M de todos los puntos (x, λ) que satisfacen $x \in U$ así como:

$$\frac{\partial L}{\partial x_j}(x, \lambda) = 0, \qquad j = 1, \ldots, n$$

$$\frac{\partial L}{\partial \lambda_i}(x, \lambda) \geq 0, \qquad \lambda_i \geq 0, \qquad \lambda_i \frac{\partial L}{\partial \lambda_i}(x, \lambda) = 0, \qquad i = 1, \ldots, l.$$

Por último evaluaremos f en cada x del conjunto $\{x \mid \text{sea } \lambda \text{ tal que } (x, \lambda) \in M\}$. El valor de x que minimiza f sobre el conjunto suele ser también un mínimo global del problema original.

Ahora supongamos que x^* es un máximo local de f en D y la cualificación de restricción se cumplen en x^*. Como x^* es factible debe satisfacer $h(x^*) \geq 0$. Según el Teorema de Kuhn y Tucker también deben existir λ^* tal que (x^*, λ^*) y satisface las condiciones del teorema y entonces (x^*, λ^*) debe ser un punto crítico de f que satisface la propiedad.

3.6. Una prueba del Teorema de Kuhn y Tucker.

Teorema de Kuhn y Tucker. *Sean* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *y* $h_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, *funciones de clase* C^1 *para* $i = 1, \ldots, l$. *Supongamos que* x^* *es un máximo local de* f *en el conjunto*

$$D = U \cap \left\{ x \in \mathbb{R}^n \mid h_i(x) \geq 0, i = 1, \ldots, l \right\},$$

donde U *es un conjunto abierto en* $\mathbb{R}^n$. *Si* $E \subset \{1, \ldots, l\}$ *denota el conjunto de restricciones en* x^* *y si* $h_E = (h_i)_{i \epsilon E}$. *Supongamos que* $\rho(Dh_E(x^*)) = |E|$, *entonces existe un vector* $\lambda^* = (\lambda_1^*, \ldots, \lambda_l^*) \in \mathbb{R}^l$ *de tal modo que se cumplen las siguientes condiciones:*

1. $\lambda_i^* \geq 0 \quad y \quad \lambda_i^* h_i(x^*) = 0 \qquad \forall i = 1 \ldots l,$

2. $Df(x^*) + \sum_{i=1}^{l} \lambda_i^* Dh_i(x^*) = 0.$

Demostración.

Sea x^* un máximo local de f en el conjunto

$$D = U \cap \{x \in \mathbb{R}^n | h_i(x) \geq 0, \ \forall i = 1, ..., l\},$$

donde cada h_i son funciones de clase C^1 y $U \subset \mathbb{R}^n$ abierto. Sea E el conjunto de restricciones efectivas en x^*, y supongamos que $\rho(Dh_E(x^*)) = |E|$, donde $h_E = (h_i)_{i \in E}$. Debemos mostrar que existe $\lambda = (\lambda_1, ..., \lambda_l) \in \mathbb{R}^l$ tal que

1. $\lambda_i \geq 0$ y $\lambda_i h_i(x^*) = 0, \quad i = 1, \ldots, l,$
2. $Df(x^*) + \sum_{i=1}^{l} \lambda_i Dh_i(x^*) = 0.$

Con la excepción de la no negatividad del vector λ, el Teorema de Kuhn y Tucker puede derivarse como consecuencia del Teorema de Lagrange. Para simplificar la notación en la prueba, denotaremos $|E|$ por k, también supongamos que las restricciones efectivas en x^* son las primeras k restricciones:

$$h_i(x^*) = 0, \quad i = 1, \ldots, k,$$
$$h_i(x^*) > 0, \quad i = k+1, \ldots, l.$$

Por supuesto, no hay pérdida de generalidad en esta suposición, ya que esto se puede lograr simplemente volviendo a numerar las restricciones.

Para cada $i \in \{1, \ldots, l\}$, definamos

$$V_i = \{x \in \mathbb{R}^n | h_i(x) > 0\}.$$

Sea $V = \bigcap_{i=k+1}^{l} V_i$. Por la continuidad de h_i, V_i es un conjunto abierto para cada i, por lo tanto V es abierto. Ahora, supongamos que $D^* \subset D$ es el conjunto de restricciones de la igualdad definido por las k primeras restricciones, es decir

$$D^* = U \bigcap \left(\bigcap_{i=k+1}^{l} V_i \right) \bigcap \{x \in \mathbb{R}^n | h_i(x) = 0, i = 1, \ldots, k\},$$

$$D^* = U \bigcap V \bigcap \{x \in \mathbb{R}^n | h_i(x) = 0, i = 1, \ldots, k\}.$$

Por construcción, tenemos $x^* \in D^*$. Dado que x^* es un máximo local de f en D, entonces es un máximo local de f en D^*. Además, tenemos $\rho(Dh_E(x*)) =$

k, por hipótesis. Por lo tanto, según el Teorema de Lagrange, existe un vector $\mu = (\mu_1, \ldots, \mu_k) \in \mathbb{R}^k$ tal que:

$$Df(x^*) + \sum_{i=1}^{k} \mu_i Dh_i(x^*) = 0.$$

Ahora definamos $\lambda \in \mathbb{R}^l$ por

$$\lambda_i = \begin{cases} \mu_i & \text{si } i = 1, \ldots, k \\ 0 & \text{si } i = k+1, \ldots, l \end{cases}$$

Mostraremos que el vector λ satisface las propiedades establecidas en el teorema.

Primero, observemos que para $i = k+1, \ldots, l$, tenemos, $\lambda_i = 0$, y por lo tanto, $\lambda_i Dh_i(x^*) = 0$. Con esto se obtiene

$$\begin{aligned} Df(x^*) + \sum_{i=1}^{l} \lambda_i Dh_i(x^*) &= Df(x^*) + \sum_{i=1}^{k} \lambda_i Dh_i(x^*) \\ &= Df(x^*) + \sum_{i=1}^{k} \mu_i Dh_i(x^*) \\ &= 0, \end{aligned}$$

que establece una de las propiedades deseadas.

Ahora, para cualquier $i \in \{\ldots, k\}$, tenemos $h_i(x^*) = 0$, así que ciertamente es el caso que $\lambda_i h_i(x^*) = 0$ para $i \in \{\ldots, k\}$. Para $i \in \{\ldots, l\}$, tenemos $\lambda_i = 0$, entonces también es el caso que $\lambda_i h_i(x^*) = 0$. Por tanto

$$\lambda_i h_i(x^*) = 0 \quad i = 1, \ldots, l.$$

Queda por demostrar que $\lambda_i \geq 0$. Como $\lambda_i = 0$ para $i = k+1, \ldots, l$, estamos obligados a mostrar que $\lambda_i \geq 0$ para $i = 1, \ldots, k$. Probaremos que $\lambda_1 \geq 0$. Con un argumento similar probaremos que $\lambda_i \geq 0$ para $i = 2, \ldots, k$.

Para este fin definamos para $x \in \mathbb{R}^n$ y $\eta \in \mathbb{R}$ la función $H : \mathbb{R}^{n+1} \to \mathbb{R}^k$, definida como $H = (H_1, \ldots, H_k)$ donde

$$\begin{aligned} H_1(x, \eta) &= h_1(x) - \eta \\ H_i(x, \eta) &= h_i(x), i = 2, \ldots, k. \end{aligned}$$

Sea $D_xH(x,\eta)$ que denota la matriz de derivada $k\times n$ de H con respecto a las variables x, y $DH_\eta(x,\eta)$ denota el vector de derivada $k\times 1$ de H con respecto a η, es decir,

$$D_xH(x,\eta)=\begin{pmatrix}\dfrac{\partial H_1}{\partial x_1}(x,\eta) & \dots & \dfrac{\partial H_1}{\partial x_n}(x,\eta)\\ \dots & \ddots & \dots\\ \dfrac{\partial H_k}{\partial x_1}(x,\eta) & \dots & \dfrac{\partial H_k}{\partial x_n}(x,\eta)\end{pmatrix}=\begin{pmatrix}\dfrac{\partial h_1(x)-\eta}{\partial x_1} & \dots & \dfrac{\partial h_1(x)-\eta}{\partial x_n}\\ \dots & \ddots & \dots\\ \dfrac{\partial h_k(x)}{\partial x_1} & \dots & \dfrac{\partial h_k(x)}{\partial x_n}\end{pmatrix},$$

$$D_\eta H(x,\eta)=\begin{pmatrix}\dfrac{\partial H_1}{\partial \eta}(x,\eta)\\ \dots\\ \dfrac{\partial H_k}{\partial \eta}(x,\eta)\end{pmatrix}=\left(\frac{\partial h_1(x)-\eta}{\partial\eta},\quad \dots,\quad \frac{\partial h_k(x)}{\partial \eta}\right)=\left(-1,0,\dots,0\right).$$

Notemos que $D_xH(x,\eta)=Dh_E(x)$, pues

$$Dh_E(x)=\begin{pmatrix}\dfrac{\partial h_1(x)}{\partial x_1} & \dots & \dfrac{\partial h_1(x)}{\partial x_n}\\ \dots & \ddots & \dots\\ \dfrac{\partial h_k(x)}{\partial x_1} & \dots & \dfrac{\partial h_k(x)}{\partial x_n}\end{pmatrix},$$

también tenemos que $DH_\eta(x,\eta)=(-1,0,...,0)$ en cualquier (x,η).

Por la definición de H, tenemos $H(x^*,0)=(0,...,0)$. Además, $\rho(D_xH(x^*,0))=\rho(Dh_E(x^*))=k$. Por lo tanto, por el Teorema de la Función implícita, hay una vecindad N del cero en $\mathbb{R}$ y $\xi:N\longrightarrow\mathbb{R}^n$ una función de clase C^1, tal que $\xi(0)=x^*$, y

$$H(\xi(\eta),\eta)=0,\eta\in N.$$

Diferenciando esta expresión respecto a η usando la regla de la cadena,

$$D_\xi H(\xi(\eta,\eta))D_\eta(\xi(\eta))+D_\eta HD_\eta\eta=0$$

y evaluando en $\xi(0)=x^*$, obtenemos

$$D_xH(x^*,0)D_\eta\xi(0)+D_\eta H(\xi(0),0)=0,$$

$$D_x H(x^*, 0) D_\eta \xi(0) + D_\eta H(x^*, 0) = 0.$$

Dado que $D_x H(x, \eta) = Dh_E(x)$ y $D_\eta H(x, \eta) = (-1, 0, \ldots, 0)$ en cualquier (x, η), esto implica

$$Dh_E(x^*) D_\eta \xi(0) = (1, 0, \ldots, 0),$$

o, de manera equivalente, que

$$DH_E(x^*) D\xi(0) = 1$$

$$DH_E(x^*) D\xi(0) = 0 \quad i = 2, \ldots, k.$$

Como $\lambda_i = 0$ para $i = k+1, \ldots, l$, ahora tenemos

$$\begin{aligned} Df(x^*) D\xi(0) &= -\left(\sum_{i=1}^{l} \lambda_i Dh_i(x^*) \right) D\xi(0) \\ &= -\left(\sum_{i=1}^{k} \lambda_i Dh_i(x^*) D\xi(0) \right) \\ &= -\lambda_1. \end{aligned}$$

Para completar la prueba, mostraremos que

$$Df(x^*) D\xi(0) \leq 0.$$

Para este fin, primero mostramos que hay $\eta^* > 0$ tal que para todo $\eta \in [0, \eta^*)$, debemos tener $\xi(\eta) \in D$, donde D es el conjunto de restricciones del problema original.

Si $\eta > 0$, como $H_i(\xi(\eta), \eta)$ para $i = 1, \ldots, k$ y de la definición de las funciones H_i, entonces

$$h_1(\xi(\eta)) = \eta > 0,$$

y

$$h_i(\xi(\eta)) = 0 \quad i = 2, \ldots, k.$$

Para $i = k+1, \ldots, l$ tenemos $h_i(\xi(0)) = h_i(x^*) > 0$. Dado que tanto h_i y ξ son continuas, se sigue que existe $\eta^* > 0$ tal que para $\eta \in [0, \eta^*)$, también podemos asegurar que

$$h_i(\xi(\eta)) > 0 \quad i = k+1, \ldots, l.$$

Finalmente, al reducir el valor de η^* si es necesario, podemos evidentemente asegurar que $\xi(\eta) \in U$ para todos $\eta \in [0, \eta^*)$. Por lo tanto, hemos encontrado $\eta^* > 0$ tal que $\xi(\eta) \in D$ para $\eta \in [0, \eta^*)$.

Ahora, dado que $\xi(0) = x^*$ es un máximo local de f en D, y $\xi(\eta)$ está en el conjunto factible para $\eta \in [0, \eta^*)$, se deduce que para $\eta > 0$ y suficientemente cerca de cero, debemos tener

$$f(\xi(\eta)) \leq f(x^*),$$

por lo tanto

$$\frac{f(\xi(\eta)) - f(x^*)}{\eta} \leq 0$$

para todos $\eta > 0$ y lo suficientemente pequeño. Hablando en términos de límite como $\eta \to 0$ obtenemos:

$$Lim_{\eta \longrightarrow 0} \frac{f(\xi(\eta)) - f(x^*)}{\eta} \leq 0,$$

de donde

$$D_\xi \left[f(\xi(\eta)) \right]\Big|_0 = Df(x^*) D\xi(0) \leq 0,$$

como queríamos.

■

3.7. Aplicaciones a la economía.

En esta sección, presentamos dos ejemplos extraídos de economía, que ilustran el uso del procedimiento para encontrar soluciones a problemas de optimización restringidos por desigualdad. El primer ejemplo, considera un problema de maximización de la utilidad. El segundo ejemplo describe un problema de minimización de costos.

Esta sección tiene un doble propósito:

1. Explicar el proceso de verificar la condición de cualificación de restricción en presencia de múltiples restricciones de desigualdad. A diferencia del caso con problemas de igualdad, donde este proceso es relativamente directo, existe la complicación adicional de que la cualificación de la restricción depende del subconjunto preciso de restricciones que son efectivas en el punto óptimo desconocido. Por lo tanto, la única manera, en general de verificar que la restricción de la cualificación se mantendrá en el punto óptimo es tomar todas las posibles asignaciones del punto y demostrar que la condición se mantendrá en cada una de estas ubicaciones.

2. Describir un método por el cual los puntos críticos de la función Lagrangena L pueden identificarse a partir de las ecuaciones que definen estos puntos. Este proceso es más complicado que la situación correspondiente en problemas de optimización con restricciones de igualdad, ya que una parte de las ecuaciones (es decir, las condiciones complementarias de exactitud) no se especifican en forma de desigualdades.

3.7.1. Aplicación a la teoría del consumidor.

En esta subsección, consideramos el problema de maximizar la función de utilidad $u(x_1, x_2) = x_1 + x_2$, donde x_1 y x_2 son las cantidades del producto 1 y 2 respectivamente y el conjunto de presupuesto está dado por

$$P(p, I) = \{(x_1, x_2) | I - p_1 x_1 - p_2 x_2 \geq 0, x_1 \geq 0, x_2 \geq 0\}$$

donde I es el ingreso, p_1 y p_2 son los precios que son términos estrictamente positivos. Notemos que hay tres restricciones de desigualdad que definen este problema:

$$h_1(x_1, x_2) = x_1 \geq 0,$$

$$h_2(x_1, x_2) = x_2 \geq 0,$$

$$h_3(x_1, x_2) = I - p_1 x_1 - p_2 x_2 \geq 0.$$

Por las condiciones anteriores se tiene $h_1 \geq 0$ y $h_2 \geq 0$.

Para resolver el problema de maximización con las restrcciones anteriores h_1, h_2 y h_3, veremos que se cumplen las condiciones del Teorema de Kuhn y

Tucker. El conjunto presupuestario $P(p, I)$ es un conjunto cerrado y acotado, es decir, compacto esto por las respectivas resticciones $h_1(x_1, x_2)$, $h_2(x_1, x_2)$ y $h_3(x_1, x_2)$, por otra parte los precios y los ingresos son todos estrictamente positivos, y la función de utilidad es continua en este conjunto. Se sigue del Teorema de Weierstrass produce la existencia de un máximo de $u(x_1, x_2)$ en P, por lo que se satisface uno de los dos requisitos del Teorema de Kuhn y Tucker.

Primero identificamos todas las combinaciones posibles de restricciones que, en principio, pueden ser efectivas, ($h_i(x) \geq 0$ es efectiva para el punto x^* si la restricción se cumple con la igualdad en x^*, es decir, si $h_i(x) = 0$) en el nivel óptimo. Como hay tres restricciones de desigualdad, hay que verificar un total de ocho combinaciones diferentes:

1. $h_1 = 0$, $h_2 = 0$ y $h_3 = 0$.
2. $h_1 = 0$, $h_2 > 0$ y $h_3 > 0$.
3. $h_1 > 0$, $h_2 = 0$ y $h_3 > 0$.
4. $h_1 > 0$, $h_2 > 0$ y $h_3 = 0$.
5. $h_1 = 0$, $h_2 = 0$ y $h_3 > 0$.
6. $h_1 = 0$, $h_2 > 0$ y $h_3 = 0$.
7. $h_1 > 0$, $h_2 = 0$ y $h_3 = 0$.
8. $h_1 > 0$, $h_2 > 0$ y $h_3 > 0$.

Dado que la función de utilidad es creciente en ambos argumentos, todos los ingresos disponibles deben agotarse en el punto óptimo, entonces tendríamos $h_3 = 0$.

Por lo tanto, solo hay tres combinaciones posibles para el conjunto h de restricciones efectivas en el óptimo, a saber, $h_E : (h_1, h_3)$ que corresponde a la combinación 6 de arriba, $h_E = (h_2, h_3)$ que corresponde a la combinación 7 de arriba, y $h_E = h_3$ que corresponde a la combinación 4 de arriba pues h_1 y h_2 no pueden ser amas cero. Mostraremos que la cualificación de restricción se cumple en cada uno de estos casos.

Si el óptimo ocurre en un punto donde la tercera restricción es efectiva, es decir, $h_E = (h_3)$ tenemos:

$$Dh_E(x_1, x_2) = (-p_1, -p_2),$$

donde el $\rho(Dh_E(x_1, x_2)) = 1$, en cualquier (x_1, x_2), además en este caso $|E| = 1$ y la cualificación de restrcción se matiene en dicho punto.

Si el óptimo ocurre en un punto donde sólo la primera y tercera restricciones son efectivas, es decir, $h_E = (h_1, h_3)$, tenemos:

$$Dh_E(x_1, x_2) = \begin{bmatrix} 1 & 0 \\ -p_1 & -p_2 \end{bmatrix},$$

en cualquier (x_1, x_2). Como p_1 y p_2 son estrictamente positivos por hipótesis, esta matriz tiene rango completo, es decir, $\rho(Dh_E(x_1, x_2)) = 2$ y como $|E| = 2$ la cualificación de restricción se mantiene.

Un cálculo similar muestra que si el óptimo ocurre en un punto donde solo las segundas y terceras restricciones son efectivas, se cumplirá la cualificación de restricción.

Por lo tanto, la condición de cualificación de restrcción del Teorema de Kuhn y Tucker también se cumple, y los puntos críticos del Lagrangiano deben contener los máximos globales del problema.

La función Lagrangiana para este problema es:

$$L(x, \lambda) = x_1 + x_2 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3(I - p_1 x_1 - p_2 x_2),$$

para encontrar el punto crítico tenemos

$$(1, 1) + \lambda_1(1, 0) + \lambda_2(0, 1) + \lambda_3(-p_1, -p_2) = (0, 0).$$

Los puntos críticos de la función Lagrangiana son las soluciones $(x_1, x_2, \lambda_1, \lambda_2, \lambda_3)$ al siguiente sistema de ecuaciones:

1. $1 + \lambda_1 - \lambda_3 p_1 = 0,$
2. $1 + \lambda_2 - \lambda_3 p_2 = 0,$

   Además se deben de cumplir las condiciones del Teorema de Kuhn y Tucker.

3. $\lambda_1 \geq 0, \quad x_1 \geq 0, \quad \lambda_1 x_1 = 0,$
4. $\lambda_2 \geq 0, \quad x_2 \geq 0, \quad \lambda_2 x_2 = 0,$
5. $\lambda_3 \geq 0, \quad I - p_1x_1 - p_2x_2 \geq 0, \quad \lambda_3(I - p_1x_1 - p_2x_2 \geq 0) = 0.$

Para resolver todos los puntos críticos de este sistema, adoptamos el siguiente procedimiento.

Consideremos un subconjunto C de $\{h_1, h_2, h_3\}$ en el que sólo las restricciones en el conjunto C se mantienen con igualdad y examinamos si hay puntos críticos . Luego, variamos C sobre todos los subconjuntos posibles de $\{h_1, h_2, h_3\}$, y así obtenemos todos los puntos críticos.

Además, como también se mencionó, el caso $C = \{h_1, h_2, h_3\}$ no tiene soluciones porque $h_1 = h_2 = 0$ implica $h_3 > 0$. Por lo tanto, solo hay que considerar tres casos: $\{h_3\}$, $\{h_2, h_3\}$ y $\{h_1, h_3\}$. Examinamos cada uno de los casos.

Caso 1: $\{h_3\}$

Ya que solo se supone que la restricción 3 se cumple con la igualdad en este caso, debemos tener $x_1 > 0$ y $x_2 > 0$. Por las condiciones de la holgura complementaria, esto implica que también debemos tener $\lambda_1 = \lambda_2 = 0$. Sustituyendo esto en las dos primeras ecuaciones que definen los puntos críticos de L, obtenemos

$$\lambda_3 p_1 = \lambda_3 p_2 = 1.$$

Lo cual implica que $p_1 = p_2$.

Si p_1 y p_2 son iguales a un valor $p > 0$, debemos tener $\lambda_3 = \frac{1}{p}$. Ahora se ve fácilmente que, cuando $p_1 = p_2 = p$, hay infinitos puntos críticos de L, que satisfacen $x_1 > 0$ y $x_2 > 0$. De hecho, cualquier punto $(x_1, x_2, \lambda_1, \lambda_2, \lambda_3)$ es un punto crítico si

$$\lambda_1 = \lambda_2 = 0, \quad \lambda_3 = \tfrac{1}{p} \quad \text{si } x_1 \in (0, 1/p) \quad \text{y} \quad x_2 = (I - px_1)/p.$$

Observemos que en todos estos puntos críticos, tenemos $u(x_1, x_2) = x_1 + x_2 = \frac{I}{p}$.

▶ En resumen: cuando h_3 es una restricción efectiva se obtiene todo un conjunto de puntos críticos

$$\left\{(x_1, x_2)|x_1 + x_2 = \frac{I}{p}, x_1 > 0, x_2 > 0\right\}.$$

Caso 2: $\{h_2, h_3\}$

Como $h_2 = h_3 = 0$ se traduce a $x_2 = 0$ y $I - p_1x_1 - p_2x_2 = 0$, debemos tener $x_1 = \frac{I}{p_1} > 0$. Por lo tanto, por la condición de holgura complementaria, debemos tener, $\lambda_1 = 0$. Por supuesto, también debemos tener $\lambda_2 \geq 0$. Substituyendo estos en las dos primeras ecuaciones que definen los puntos críticos de L, obtenemos

$$\lambda_3 p_1 = 1 \leq 1 + \lambda_2 = \lambda_3 p_2.$$

Por lo tanto, ese punto crítico sólo puede existir si $p_1 \leq p_2$, además se obtiene $\lambda_3 = \dfrac{1}{p_1}$, de donde $\lambda_2 = \dfrac{p_2}{p_1} - 1$. Por lo tanto el único punto crítico de L en el que $h_2 = h_3 = 0$ es

$$(x_1, x_2, \lambda_1, \lambda_2, \lambda_3) = \left(\frac{I}{p_1}, 0, 0, \frac{p_2}{p_1} - 1, \frac{1}{p_1}\right).$$

El valor de la función objetivo en este punto crítico viene dado por $u(x_1, x_2) = \frac{I}{p_1}$.

▶ L tiene sólo un único punto crítico en $(x_1, x_2) = (\dfrac{I}{p_1}, 0)$. Además $p_1 \leq p_2$.

Caso 3: $\{h_1, h_3\}$

Esto es similar al Caso 2. Es un punto crítico de L sólo si $p_1 \geq p_2$. El único punto crítico de L en este caso está dado por:

$$(x_1, x_2, \lambda_1, \lambda_2, \lambda_3) = \left(0, \frac{I}{p_2}, \frac{p_1}{p_2} - 1, 0, \frac{1}{p_2}\right)$$

El valor de la función objetivo en este punto crítico viene dado por $u(x_1, x_2) = \frac{I}{p_2}$.

▶ Hay exactamente un punto crítico de L en $(x_1, x_2) = (0, \frac{I}{p_2})$. En este caso $p_1 \geq p_2$.

Los puntos críticos de L deben contener la solución del problema. Como L tiene un punto crítico único cuando $p_1 \geq p_2$, se deduce que este punto crítico es el máximo global del problema para esta configuración de parámetros; es decir, la solución única al problema cuando $p_1 \geq p_2$ es $(x_1, x_2) = (0, \frac{I}{p_2})$. Del mismo modo, la única solución al problema cuando $p_1 \leq p_2$ es $(x_1, x_2) = (\frac{I}{p_1}, 0)$. Finalmente, cuando $p_1 = p_2 = p$, hay infinitos puntos críticos de L, pero en todos estos tenemos $u(x_1, x_2) = \frac{I}{p}$.

Por lo tanto, cada uno de estos valores de (x_1, x_2) define un máximo global del problema en este caso.

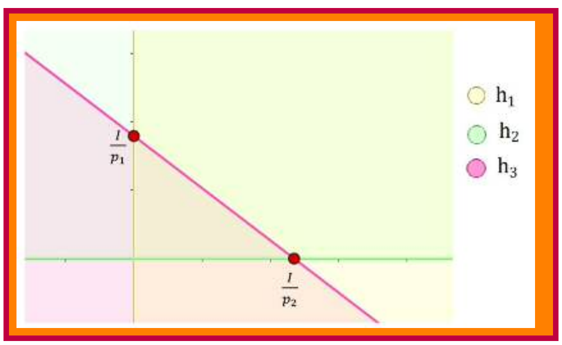

Figura 3.4: Representación geométrica del conjunto de restricciones para el ejemplo de la teoría del consumidor.

Para ver más ejemplos relacionados consultar [1], [9] y [10].

3.7.2. Aplicaciones a la teoría del productor

El problema que consideramos en esta sección es el de minimizar $f(x_1, x_2) = w_1x_1 + w_2x_2$, donde x_1 representa la cantidad del producto 1 y x_2 la cantidad del producto 2 y w_1 y w_2 los costos de producción de cada producto respectivamente, con un costo presupuesto $y > 0$, sobre el conjunto factible

$$D = \left\{(x_1, x_2) \in \mathbb{R}^2_+ | x_1^2 + x_2^2 \geq y\right\}.$$

El conjunto de restricciones de este problema se define a través de tres restricciones de desigualdad, a saber

$$h_1(x_1, x_2) = x_1 \geq 0,$$

$$h_2(x_1, x_2) = x_2 \geq 0,$$

$$h_3(x_1, x_2) = x_1^2 + x_2^2 - y \geq 0.$$

Como el problema anterior, supondremos que todos los parámetros del problema, w_1, w_2 y y son estrictamente positivos.

Una vez más, comenzamos nuestro análisis con una demostración de que ambas condiciones del Teorema de Kuhn y Tucker se cumplen.

Para comprobar que la cualificación de la restricción se cumple, primero identificamos todas las combinaciones posibles de las restricciones que, en principio, pueden mantener la igualdad en el nivel óptimo.

Como hay tres restricciones de desigualdad, hay que verificar un total de ocho combinaciones diferentes:

1. $h_1 = 0$, $h_2 = 0$ y $h_3 = 0$.
2. $h_1 = 0$, $h_2 > 0$ y $h_3 > 0$.
3. $h_1 > 0$, $h_2 = 0$ y $h_3 > 0$.
4. $h_1 > 0$, $h_2 > 0$ y $h_3 = 0$.
5. $h_1 = 0$, $h_2 = 0$ y $h_3 > 0$.
6. $h_1 = 0$, $h_2 > 0$ y $h_3 = 0$.
7. $h_1 > 0$, $h_2 = 0$ y $h_3 = 0$.
8. $h_1 > 0$, $h_2 > 0$ y $h_3 > 0$.

De estos, el primero puede descartarse ya que $h_1 = h_2 = 0$ implica, $x_1^2 + x_2^2 = 0$, mientras que el conjunto de restricciones requiere $x_1^2 + x_2^2 \geq y$. También es evidente que, como w_1 y w_2 son estrictamente positivos, debemos tener $h_3 = 0$

en el valor óptimo (es decir, la producción total $x_1^2 + x_2^2$ debe ser exactamente igual a y), o los costos podrían reducirse disminuyendo la producción.

Esto significa que solo hay tres posibles descripciones del conjunto h_E de restricciones efectivas en el óptimo: $h_E = h_3, h_E = (h_1, h_3)$ y $h_E = (h_2, h_3)$. Mostraremos que en cada caso, la cualificación de restricción se mantiene. Primero, considere el caso $h_E = (h_1, h_3)$. Como h_1 y h_3 son efectivos, tenemos $x_1 = 0$ y $x_1^2 + x_2^2 = y$, entonces $x_2 = \sqrt{y}$. Por lo tanto,

$$Dh_E(x_1, x_2) = \begin{bmatrix} 1 & 0 \\ 2x_1 & 2x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2\sqrt{y} \end{bmatrix}.$$

Así el $\rho Dh_E = 2$. Como se cumple la cualificación de restricción, se sigue que el óptimo ocurre en un punto donde h_1 y h_3 son las únicas restricciones efectivas.

El caso $h_E = (h_2, h_3)$ es análogo. Para el caso $h_E = h_3$ tenemos

$$Dh_E(x_1, x_2) = (2x_1, 2x_2),$$

de donde $\rho Dh_E = 1$.

En resumen, el óptimo debe ocurrir en un punto donde $h_3 = 0$, pero no importa en qué parte de este conjunto se produzca el óptimo, se cumplirá la cualificación de restricción. Se deduce que el conjunto de puntos críticos del Lagrangiano debe contener la(s) solución(es) del problema.

El Lagrangiano L en este problema tiene la forma:

$$L(x_1, x_2, \lambda_1, \lambda_2, \lambda_3) = -w_1 x_1 - w_2 x_2 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3(x_1^2 + x_2^2 - y).$$

Utilizando el Teorema de Lagrange para encontrar el punto crítico tenemos

$$\begin{aligned} Df(x_1, x_2) + \sum_{i=1}^{l} \lambda_i Dh_i(x_1, x_2) &= 0, \\ (-w_1, -w_2) + \lambda_1(1, 0) + \lambda_2(0, 1) + \lambda_3(2x_1, 2x_2) &= (0, 0). \end{aligned}$$

Hemos establecido implícitamente el problema como un problema de maximización, con la función objetivo $-w_1 x_1 - w_2 x_2$. Los puntos críticos de L son las soluciones $(x_1, x_2, \lambda_1, \lambda_2, \lambda_3)$ al siguiente sistema de ecuaciones:

1. $-w_1 + \lambda_1 + 2\lambda_3 x_1 = 0$
2. $-w_2 + \lambda_2 + 2\lambda_3 x_2 = 0$.

   Además se deben de cumplir las condiciones del Teorema de Kuhn y Tucker.
3. $\lambda_1 \geq 0, \quad x_1 \geq 0, \quad \lambda_1 x_1 = 0$
4. $\lambda_2 \geq 0, \quad x_2 \geq 0, \quad \lambda_2 x_2 = 0$
5. $\lambda_3 \geq 0, \quad x_1^2 + x_2^2 - y \geq 0, \quad \lambda_3(x_1^2 + x_2^2 - y) = 0$.

Ahora tomamos los subconjuntos C de $\{h_1, h_2, h_3\}$ y buscaremos el conjunto de todas las soluciones posibles para las ecuaciones anteriores.

Como hemos mencionado, h_3 debe ser eficaz a un nivel óptimo, por lo que basta con encontrar todos los puntos críticos de L a los que $h_3 = 0$. Se descarta el caso $C = \{h_1, h_2, h_3\}$, ya que $h_1 = h_2 = 0$ viola la tercera restricción que es $h_3 \geq 0$. Esto da como resultado los posibles subconjuntos, $C = \{h_3\}$, $C = \{h_2, h_3\}$ y $C = \{h_1, h_3\}$. Analizaremos cada uno de estos a su vez.

Caso 1: $C = \{h_3\}$

Como debemos tener $x_1 > 0$ y $x_2 > 0$ en este caso, las condiciones de holgura complementarias implican $\lambda_1 = \lambda_2 = 0$. Sustituyendo estos valores en las primeras dos ecuaciones que definen los puntos críticos de L, obtenemos

$$2\lambda_3 x_1 = w_1,$$

$$2\lambda_3 x_2 = w_2.$$

Dividiendo la primera ecuación por la segunda, obtenemos

$$x_1 = \left(\frac{w_1}{w_2}\right) x_2.$$

Por hipótesis, también tenemos $h_3 = 0$ o, $x_1^2 + x_2^2 = y$. Por lo tanto, tenemos

$$x_1 = \left(\frac{w_1^2 y}{w_1^2 + w_2^2}\right)^{1/2}, \qquad x_2 = \left(\frac{w_2^2 y}{w_1^2 + w_2^2}\right)^{1/2}, \qquad \lambda_3 = \left(\frac{w_1^2 + w_2^2}{y}\right)^{1/2}$$

Combinado con $\lambda_1 = \lambda_2 = 0$, esto representa el único punto crítico de L en el que $h_1 > 0$, $h_2 > 0$, y $h_3 = 0$. El valor de la función objetivo $-w_1x_1 - w_2x_2$ en el punto crítico es

$$-w_1x_1 - w_2x_2 = -(w_1^2 + w_2^2)^{1/2}y^{1/2}.$$

Caso 2: $C = \{h_2, h_3\}$

En este caso, tenemos $x_1 > 0$ y de hecho, $x_1 = \sqrt{y}$, ya que si $h_2 = h_3 = 0$ entonces $x_2 = 0$ y $x_1^2 + x_2^2 - y = 0$. Por lo tanto, por las condiciones de holgadura complementaria también debemos tener $\lambda_1 = 0$. Sustituyendo lo anterior en la primera de las ecuaciones que definen los puntos críticos de L, obtenemos

$$2\lambda_3x_1 - w_1 = 2\lambda_3\sqrt{y} - w_1 = 0,$$

de donde $\lambda_3 = w_1/2\sqrt{y}$. Finalmente, sustituyendo $x_2 = 0$ en la segunda ecuación que define los puntos críticos de L, obtenemos $\lambda_2 = w_2$. Por lo tanto, el único punto crítico de L que satisface $h_2 = h_3 = 0$ es

$$(x_1, x_2, \lambda_1, \lambda_2, \lambda_3) = (\sqrt{y}, 0, 0, w_2, w_1\ 2/\sqrt{y}).$$

El valor de la función objetivo $(-w_1x_1 - w_2x_2)$ en este crítico es

$$-w_1x_1 - w_2x_2 = -w_1y^{1/2}.$$

Caso 3: $C = \{h_1, h_3\}$.

Esto es lo mismo que el Caso 2, con los cambios obvios. El único punto crítico de L que cae en este caso es

$$(x_1, x_2, \lambda_1, \lambda_2, \lambda_3) = (0, \sqrt{y}, w_1, 0, w_2/2\sqrt{y}),$$

y el valor de la función objetivo $(-w_1x_1 - w_2x_2)$ en este punto crítico es

$$(-w_1x_1 - w_2x_2) = -w_2y^{1/2}$$

Obtuvimos que, L tiene tres puntos críticos, y los valores que toma la función objetivo en estos tres puntos son $-(w_1^2+w_2^2)^{1/2}y^{1/2}$, $-w_1y^{1/2}$ y $-w_2y^{1/2}$. Ahora solo resta comparar el valor de la función objetivo en los tres puntos críticos.

Dado que $w_1 > 0$ y $w_2 > 0$, siempre es el caso que $(w_1^2 - w_2^2)^{1/2} > w_1$, por lo tanto, que

$$-(w_1^2 + w_2^2)^{1/2} y^{1/2} < -w_1 y^{1/2}.$$

Como consecuencia, podemos descartar el primer valor de la función objetivo y el punto que representa. Comparando el valor de la función objetivo en los dos puntos restantes, se puede ver que

- Cuando $w_1 < w_2$, entonces el mayor de los dos valores es $-w_1 y^{1/2}$, que lo alcanza en $(x_1, x_2) = (y^{1/2}, 0)$. Por lo tanto, el problema tiene una solución única cuando $w_1 < w_2$, es decir $(y^{1/2}, 0)$.
- Cuando $w_1 > w_2$, entonces el mayor de los dos valores es $-w_1 y^{1/2}$, que surge en $(x_1, x_2) = (0, y^{1/2})$. Por lo tanto, el problema tiene una solución única cuando $w_1 > w_2$, es decir $(0, y^{1/2})$.
- Cuando w_1 y w_2 tienen un valor común $w > 0$, entonces estos dos valores de la función objetivo coinciden. Por lo tanto, si $w_1 = w_2$, el problema tiene dos soluciones, concretamente $(y^{1/2}, 0)$ y $(0, y^{1/2})$.

Conclusión

Del ejemplo se tiene que los puntos donde la función alcanza su mínimo están dados en función del valor constante y, es decir, el mínimo costo de producción depende del presupuesto inicial y.

Para ver más ejemplos relacionados consultar [7], [9] y [10].

Capítulo 4

Programación Convexa.

La programación convexa (PC) se refiere a la minimización de una función convexa de una o varias variables en un conjunto convexo. El conjunto convexo a menudo se define en términos de desigualdades que involucran otras funciones convexas.

Sean $f : C \to \mathbb{R}$ y $g_i : C \to \mathbb{R}$ con $i = 1, ..., l$, funciones convexas definidas en un subconjunto convexo cerrado no vacío C de $\mathbb{R}^l$.

El problema principal en la programación convexa es el siguiente:

$$\text{minimizar } f(x) \text{ sujeto a condiciones } g_i(x) \leq 0 \text{ para } i = 1, ..., l. \qquad \text{(P)}$$

Para mayor comodidad de notación, definamos $g(x) = (g_1(x), ..., g_l(x))$, denotemos $g(x) \leq 0$, si $g_i(x) \leq 0$ para toda $i = 1, ..., l$, entonces el problema es equivalente a

$$\text{minimizar } f(x) \text{ sujeto a la condición } g(x) \leq 0.$$

***Definición* 4.1.** *El conjunto factible para el problema principal, es el conjunto*

$$\begin{aligned} F &= \{x \in C | g(x) \leq 0\} \\ &= \bigcap_{i=1}^{l} \{x | g_i(x) \leq 0\}. \end{aligned}$$

Los elementos de F serán llamados puntos factibles.

***Definición* 4.2.** *Se dice que el problema es **consistente** si F es no vacío, y es **superconsistente** si hay x en F con $g_i(x) < 0$ para todo $i = 1, ..., l$, tal punto se llama un punto Slater.*

***Definición* 4.3.** *Para cada z en $\mathbb{R}^l$ sea*

$$MP(z) = inf \{f(x)|x \in C, g(x) \leq z\},$$

en particular denotemos $MP = MP(0)$.

Recordemos que la desigualdad $g(x) \leq z$ denota que se cumple $g_i(x) \leq z$ para cada $i = 1, ..., l$. También denotaremos como $P(z)$ al problema de minimizar la función $f(x)$ para x en C con la restricción $g(x) \leq z$, con esta notación $P = P(0)$, llamaremos a $P(z)$ **el problema perturbado**.

El conjunto factible para $P(z)$ es

$$F_z = \{x|g(x) \leq z\}.$$

Analizaremos las propiedades de la función $MP(z)$, en particular, cómo se comporta la función $MP(z)$ cuando z tiende a 0.

El siguiente ejemplo muestra la relación de $MP(z)$ cuando z tiende a 0.

Ejemplo. *Sea $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x^2$, sabemos que el mínimo ocurre en $x = 0$, y $f(0) = 0$ es el valor mínimo global.*

Consideremos el problema perturbado, es decir, minimizar $f(x) = x^2$ restringido a $x \leq z$. Si $z < 0$,como $x \leq z$, entonces $f(z) \leq f(x)$ así $f(z)$ es mínimo local. Para $z \geq 0$, el mínimo del problema perturbado es el mínimo global, que está en $x = 0$, se concluye que $MP(z) = 0$. El mínimo global de $MP(z)$ también ocurre en $z = 0$.

***Teorema* 4.1.** *Si P es un problema convexo y si $P(z)$ es la perturbación de P por $z \in \mathbb{R}^l$, entonces la función $MP(z)$ es convexa y su dominio es un subconjunto convexo de $\mathbb{R}^l$. Si P es superconsistente, entonces $z = 0$ es un punto interior del dominio de $MP(z)$.*

Demostración.

Como el dominio de MP es el conjunto de puntos $z \in \mathbb{R}^l$ para los cuales $F_z \neq \emptyset$. Sean $z^{'}$, $z^{''}$ que pertenecen al dominio de $MP(z)$ y $0 \leq \lambda \leq 1$, Notemos que $\lambda x^{'} + (1-\lambda)x^{''}$ es factible para $P(\lambda z^{'} + (1-\lambda)z^{''})$, además $x^{'}$, $x^{''}$ son factibles para $P(z^{'})$ y $P(z^{''})$ respectivamente, en efecto

$$\begin{aligned} g(\lambda x^{'} + (1-\lambda)x^{''}) &\leq \lambda g(x') + (1-\lambda)g(x'') \\ &\leq \lambda z^{'} + (1-\lambda)z^{''}. \end{aligned}$$

Por lo tanto el dominio de $MP(z)$ es un subconjunto convexo de $\mathbb{R}^m$. Ahora veamos que MP es una función convexa, supongamos que $MP(z)$ es finita, entonces

$$\begin{aligned} MP(\lambda z' + (1-\lambda)z'') &= inf\left\{f(x) : x \in C,\, g(x) \leq \lambda z' + (1-\lambda)z''\right\} \\ &= inf\left\{f(x) : x = \lambda x' + (1-\lambda)x'' \text{ donde } x', x'' \in C \text{ y } g(x) \leq \lambda z' + (1-\lambda)z''\right\} \\ &\leq inf\left\{f(x) : x = \lambda x' + (1-\lambda)x'' \text{ donde } x', x'' \in C \text{ y } \lambda g(x') + (1-\lambda)g(x'') \leq \lambda z' + (1-\lambda z^{''})\right\} \\ &\leq inf\left\{f(\lambda x' + (1-\lambda)x'') : x', x'' \in C \text{ y } g(x') \leq z' \text{ , } g(x'') \leq z''\right\} \\ &\leq inf\left\{\lambda f(x') + (1-\lambda)f(x'') : x', x'' \in C \text{ y } g(x') \leq z' \text{ , } g(x'') \leq z''\right\} \\ &\leq inf\left\{\lambda f(x') : x' \in C \text{ y } g(x') \leq z'\right\} + inf\left\{(1-\lambda)f(x'') : x'' \in C \text{ y } g(x'') \leq z''\right\} \\ &= \lambda MP(z') + (1-\lambda)MP(z''). \end{aligned}$$

Por otro lado, si $MP(\lambda z^{'} + (1-\lambda)z^{''}) = -\infty$ entonces claramente $MP(\lambda z^{'} + (1-\lambda)z^{''}) \leq \lambda MP(z^{'}) + (1-\lambda)MP(z^{''})$ en consecuencia $MP(z)$ es convexa en su dominio.

Ahora supongamos que P es superconsistente, es decir, existe un $w \in \mathbb{R}^n$, tal que $g_i(w) < 0$ con $i = 1, ..., l$. Sea

$$r = min\left\{-g_i(w) : 1 \leq i \leq l\right\} > 0.$$

Para la bola abierta con radio r y centro 0, $B_r(0) \subset \mathbb{R}^l$, podemos encontrar $z = (z_1, ..., z_l)$ tal que $-r < z_i < r$ para todo i, de modo que

$$g_i(w) \leq -r < z_i \quad i = 1, ..., l.$$

Por lo tanto, $P(z)$ es un problema consistente para todas las $B_r(0)$, esto demuestra que es un punto interior del dominio de $MP(z)$.

$\blacksquare$

Por el Teorema 1.8, si el problema es superconsistente, entonces hay un vector u tal que

$$MP(z) \geq MP(0) + \langle u, z - 0 \rangle.$$

De hecho si el problema es superconsistente, podemos demostrar que $u \leq 0$. Supongamos sin pérdida de generalidad que existe $i_0 \in \{1, ..., l\}$ tal que $u_{i_0} > 0$ y $u_i \leq 0$ para todo $i \neq i_0$. Como $z = 0$ está en el interior del dominio de $MP(z)$, existe $r > 0$, tal que F_z es no vacío para toda z con $||z|| < r$ en particular para $z = 0$. Sea $w_j = 0$ para $j \neq i_0$ y $w_{i_0} = r/2$, entonces F_w es no vacío y $F_0 \subseteq F_w$, de la definición de MP se tiene $MP(0) \geq MP(w)$. De la ecuación

$$\begin{aligned} MP(z) &\geq MP(0) + \langle u, z - 0 \rangle \\ &= MP(0) + (u_1 w_1 +, ..., + u_l w_l) \end{aligned}$$

tenemos $MP(w) \geq MP(0) + \dfrac{r}{2} u_{i_0} > MP(0)$. Esto es una contradicción, por lo que $u \leq 0$.

La ecuación

$$MP(z) \geq MP(0) + \langle u, z - 0 \rangle,$$

la escribimos como

$$-\langle u, z - 0 \rangle \geq MP(0) - MP(z).$$

Si ahora denotamos $\lambda^* = -u$ en lugar de u. Para cualquier z tenemos

$$\langle \lambda^*, z \rangle \geq MP(0) - MP(z),$$

para $z \geq 0$ tenemos $MP(z) \leq MP(0)$, y

$$\langle \lambda^*, z \rangle \geq MP(0) - MP(z) \geq 0.$$

El número $\langle \lambda^*, z \rangle$ mide la razón de cambio de $MP(z)$ a medida que aumentamos z lejos de $z = 0$; por esta razón, λ^* se llama el **vector de sensibilidad**, así como el vector de los multiplicadores de Lagrange.

***Definición* 4.4.** *Definamos la función de Lagrange para el problema P*

$$L(x, \lambda) = f(x) + \sum_{i=1}^{l} \lambda_i g_i(x) = f(x) + \langle \lambda, g(x) \rangle$$

para todas las x en C y $\lambda \geq 0$.

***Definición* 4.5.** *Para cada x fijo en C, definimos*

$$F(x) = sup_{\lambda \geq 0} L(x, \lambda).$$

Si x es factible para P, entonces

$$\bigcap_{i=1}^{l} \{x \in C : g_i(x) \leq 0\} \neq \emptyset \text{ y } \lambda_i g_i(x) \leq 0 \text{ para } i = 1, ..., l,$$

por lo que $f(x) \geq L(x, \lambda)$ para todo $\lambda \geq 0$, de modo que $f(x) \geq F(x)$.

Como $f(x) = L(x, 0)$ y por definición de $F(x)$, se tiene que $L(x, 0) \leq F(x)$, se tiene que $f(x) = F(x)$ para todo punto factible x en C. Si x no es factible en C, entonces $\lambda_i g_i(x) > 0$ de donde $F(x)$ no es acotado para valores adecuados de λ por tanto $F(x) = +\infty$.

De lo anterior se deduce que, minimizar $f(x)$ en un conjunto factible en C es equivalente a minimizar $F(x)$ sobre todas las x en C, es decir, se elimina la restricción de que x sea factible para P.

***Teorema* 4.2.** *Si el problema P tiene un vector de sensibilidad $\lambda^* \geq 0$ y es superconsistente, entonces*

$$MP(0) = inf_{x \in C} L(x, \lambda^*).$$

Demostración.

Para cualquier x fijo y $t \in C$, consideremos el conjunto factible

$$F_{g(x)} = \{t | g(t) \leq g(x)\},$$

si $t = x$ se tiene $g(x) \leq g(x)$, es decir, $x \in F_{g(x)}$. Recordemos que si P es superconsistente entonces $\langle \lambda^*, z \rangle \geq MP(0) - MP(z) \geq 0$ para cada $z \in \mathbb{R}$, en particular para $z = g(x)$,

$$MP(g(x)) + \langle \lambda^*, g(x) \rangle \geq MP(0).$$

Como $x \in F_{g(x)}$, tenemos

$$f(x) \geq MP(g(x)),$$

de donde

$$f(x) + \langle \lambda^*, g(x) \rangle \geq MP(0).$$

Por lo tanto,

$$inf_{x \in C} \left(f(x) + \langle \lambda^*, g(x) \rangle \right) \geq MP(0). \tag{1}$$

Por otro lado

$$inf_{x \in C} \left(f(x) + \langle \lambda^*, g(x) \rangle \right) \leq inf_{x \in C, g(x) \leq 0} \left(f(x) + \langle \lambda^*, g(x) \rangle \right)$$

y

$$inf_{x \in C, g(x) \leq 0} \left(f(x) + \langle \lambda^*, g(x) \rangle \right) \leq inf_{x \in C, g(x) \leq 0} f(x) = MP(0)$$

que junto con (1) se obtiene el resultado.

■

Para mayor información ver [3], [5] y [6].

4.1. El Teorema de Karush-Kuhn-Tucker.

Comenzamos con condiciones suficientes para que un vector x^* pueda ser una solución al problema de Programación Convexa (PC). Bajo ciertas restricciones, como lo especifica el Teorema de Karush-Kuhn-Tucker, estas condiciones también son necesarias. Seguiremos denotando $C \subset \mathbb{R}^n$ al conjunto convexo.

***Teorema* 4.3.** *Sea x^* un elemento de C. Si existe $\lambda^* \geq 0$ tal que, para cada $x \in C$ y cada vector $\lambda \geq 0$, se cumple*

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$$

entonces x^* es factible y x^* resuelve el problema de PC.

Demostración.

Como $L(x^*, \lambda) \leq L(x^*, \lambda^*)$ entonces por definición de función de Lagrange se tiene

$$\begin{aligned} f(x^*) + \sum_{i=1}^{l} \lambda_i g_i(x^*) &\leq f(x^*) + \sum_{i=1}^{l} \lambda_i^* g_i(x^*) \\ \sum_{i=1}^{l} \lambda_i g_i(x^*) &\leq \sum_{i=1}^{l} \lambda_i^* g_i(x^*), \end{aligned}$$

es decir,

$$\begin{aligned}\sum_{i=1}^{l}\lambda_i g_i(x^*) &- \sum_{i=1}^{l}\lambda_i^* g_i(x^*) \leq 0,\\ \sum_{i=1}^{l}(\lambda_i-\lambda_i^*)g_i(x^*) &\leq 0. \text{ para cada } \lambda \geq 0.\end{aligned}$$

Sea $j \in \{1,...,l\}$ fijo, definamos

$$\lambda_i = \begin{cases} \lambda_i^* & \text{si } i \neq j. \\ \lambda_i^*+1 & \text{si } i = j. \end{cases}$$

Entonces $\lambda_i^* \geq 0$. Como

$$\begin{aligned}\sum_{i=1}^{l}(\lambda_i-\lambda_i^*)g_i(x^*) &= \sum_{i=1}^{j}-1(\lambda_i-\lambda_i^*)g_i(x^*)+(\lambda_j-\lambda_j^*)g_i(x^*)+\sum_{i=j+1}^{l}(\lambda_i-\lambda_i^*)g_i(x^*)\\ &= (\lambda_i-\lambda_i^*)g_i(x^*)\\ &= (\lambda_i^*+1-\lambda_i^*)g_i(x^*)\end{aligned}$$

por lo tanto $g_i(x^*) \leq 0$ de modo que x^* es factible para P.

Por otro lado, como

$$L(x^*,\lambda) \leq L(x^*,\lambda^*)$$

esto también implica

$$f(x^*) = L(x^*,0) \leq L(x^*,\lambda^*),$$

es decir

$$f(x^*) \leq L(x^*,\lambda^*) = f(x^*) + \sum_{i=1}^{l}\lambda_i^* g_i(x^*),$$

pero $\lambda^* \geq 0$ y $g_i(x^*) \leq 0$ para $i = 1,...,l$ de modo que

$$\begin{aligned} f(x^*) + \sum_{i=1}^{l} \lambda_i^* g_i(x^*) &\leq f(x^*), \\ L(x^*, \lambda^*) &\leq f(x^*) \end{aligned}$$

por lo que

$$f(x^*) = L(x^*, \lambda^*)$$

Aplicando propiedades de ínfimos

$$\begin{aligned} inf_{x \in C} f(x^*) = inf_{x \in C} L(x^*, \lambda^*) &\leq inf \left\{ L(x, \lambda^*) : x \in C \right\}, \\ &\leq inf \left\{ L(x, \lambda^*) : x \in C, g_1 \leq 0, ..., g_l \leq 0 \right\}, \\ &= inf \left\{ f(x) + \sum_{i=1}^{l} \lambda_i^* g_i(x) : x \in C, g_1 \leq 0, ..., g_l \leq 0 \right\}, \\ &\leq inf \left\{ f(x) : x \in C, g_1 \leq 0, ..., g_l \leq 0 \right\}, \\ = MP. \end{aligned}$$

Por lo tanto $f(x^*) = MP$.

Así x^* resuelve el problema

■

***Corolario* 4.1.** *Sea $x^* \in C$ y $\lambda^* \geq 0$ tal que $L(x^*, \lambda^*) \leq L(x, \lambda^*)$ para todo $x \in C$ y $\lambda_i^* g_i(x^*) = 0$ para todo $i = 1, ..., l$, entonces x^* es factible y x^* resuelve el problema de PC.*

Demostración.

Como $L(x^*, \lambda^*) \leq L(x, \lambda^*)$ aplicando la definición de función de Lagrange se tiene

$$f(x^*) + \sum_{i=1}^{l} \lambda_i^* g_i(x^*) \leq f(x) + \sum_{i=1}^{l} \lambda_i^* g_i(x),$$

notemos que $\lambda_i^* g_i(x^*) = 0$ para todo $i = 1, ..., l$, aplicando propiedades de los ínfimos se tiene

$$inf_{x \in C} f(x^*) = L(x^*, \lambda^*) = inf_{x \in C} L(x, \lambda^*),$$

o bien

$$
\begin{aligned}
f(x^*) = L(x^*, \lambda^*) &\leq inf\left\{L(x,\lambda^*) : x \in C\right\}, \\
&\leq inf\left\{L(x,\lambda^*) : x \in C, g_1 \leq 0, ..., g_l \leq 0\right\}, \\
&= inf\left\{f(x) + \sum_{i=1}^{l} \lambda_i^* g_i(x) : x \in C, g_1 \leq 0, ..., g_l \leq 0\right\}, \\
&\leq inf\left\{f(x) : x \in C, g_1 \leq 0, ..., g_l \leq 0\right\}, \\
= MP.
\end{aligned}
$$

Por lo tanto x^* es un punto factible para P tal que $f(x^*) = MP$, es decir, x^* es una solución para P.

■

4.2. Algunas versiones del Teorema de Karush-Kuhn-Tucker.

4.2.1. La forma de punto silla del teorema.

Esta forma del Teorema de Karush-Kuhn-Tucker (KKT) no requiere que las funciones involucradas sean diferenciables. La forma de silla de montar del Teorema de KKT es la siguiente.

***Teorema* 4.4.** *Sea P el problema principal de PC superconsistente. Entonces x^* resuelve P si y sólo si hay un vector λ^* tal que*

1. $\lambda^* \geq 0$.
2. $L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$ *para todos los $x \in C$ y todos los $\lambda \geq 0$.*
3. $\lambda_i^* g_i(x^*) = 0$, *para todo $i = 1, ..., l$.*

Demostración.

Dado que P es superconsistente y x^* resuelve P, sabemos por el Teorema 4.2 que hay $\lambda^* \geq 0$ tal que

$$f(x^*) = inf_{x \in C} L(x, \lambda^*).$$

Como

$$f(x^*) \leq L(x^*, \lambda^*) = f(x^*) + \langle \lambda^*, g(x^*) \rangle$$

y dado que $\lambda^* \geq 0$ y $g(x^*) \leq 0$, también tenemos

$$f(x^*) + \langle \lambda^*, g(x^*) \rangle \leq f(x^*).$$

Ahora podemos concluir que $f(x^*) = L(x^*, \lambda^*)$ y $\langle \lambda^*, g(x^*) \rangle = 0$. Se deduce que $\lambda_i^* g_i(x^*) = 0$, para todos $i = 1, ..., l$. Entonces, para $\lambda \geq 0$,

$$\begin{aligned} L(x^*, \lambda^*) - L(x^*, \lambda) &= f(x^*) + \sum_{i=1}^{l} \lambda_i^* g_i(x^*) - \left(f(x^*) + \sum_{i=1}^{l} \lambda_i g_i(x^*) \right) \\ &= \sum_{i=1}^{l} \lambda_i^* g_i(x^*) - \sum_{i=1}^{l} \lambda_i g_i(x^*) \\ &= - \sum_{i=1}^{l} \lambda_i g_i(x^*) \end{aligned}$$

también tenemos

$$L(x^*, \lambda) \leq L(x^*, \lambda^*),$$

para todos $\lambda \geq 0$.

A la inversa, supongamos que x^* y λ^* satisfacen las tres condiciones del teorema. Primero, mostramos que x^* es factible para P, es decir, $g(x^*) \leq 0$. De la misma forma que el teorema anterior, definamos un $j \in \{1, ..., l\}$ fijo, tal que

$$\lambda_i = \begin{cases} \lambda_i^* & \text{si } i \neq j. \\ \lambda_i^* + 1 & \text{si } i = j. \end{cases}$$

Entonces nosotros tenemos $\lambda \geq 0$.

Del punto 2 de la hipótesis

$$
\begin{aligned}
0 \leq L(x^*,\lambda^*) - L(x^*,\lambda) &= f(x^*) + \sum_{i=1}^{l} \lambda_i^* g_i(x^*) - \left(f(x^*) + \sum_{i=1}^{l} \lambda_i g_i(x^*) \right) \\
&= \sum_{i=1}^{l} \lambda_i^* g_i(x^*) - \sum_{i=1}^{l} \lambda_i g_i(x^*) \\
&= \sum_{i=1}^{l} (\lambda_i^* - \lambda_i)\, g_i(x^*) \\
&= \sum_{i=1}^{j-1} (\lambda_i^* - \lambda_i)\, g_i(x^*) + \left(\lambda_j^* - \lambda_j\right) g_i(x^*) + \sum_{i=j+1}^{l} (\lambda_i^* - \lambda_i)\, g_i(x^*) \\
&= (\lambda_i^* - \lambda_i)\, g_i(x^*) \\
&= (\lambda_i^* - (\lambda_i^* + 1))\, g_i(x^*) \\
&= -g_i(x^*)
\end{aligned}
$$

Por definición de la función Lagrangeana sabemos que $L(x^*,\lambda) = f(x^*) + \langle \lambda^*, g(x^*) \rangle$, haciendo $\lambda = 0$, tenemos

$$
\begin{aligned}
f(x^*) &= L(x^*,0) \\
&\leq L(x^*,\lambda^*) = f(x^*) + \langle \lambda^*, g(x^*) \rangle
\end{aligned}
$$

pero $\langle \lambda^*, g(x^*) \rangle = 0$, por lo que

$$
f(x^*) = L(x^*,0) \leq L(x^*,\lambda^*) = f(x^*)
$$

así que

$$
f(x^*) = L(x^*,\lambda^*) \leq L(x,\lambda^*), \tag{4}
$$

pero por el corolario 4.1 también tenemos

$$
L(x^*,\lambda^*) \leq inf_{x\in C} \left(f(x) + \langle \lambda^*, g(x) \rangle \right) \leq inf_{x\in C, g(x)\leq 0} f(x) = MP(0). \tag{5}
$$

Por lo tanto por las ecuaciones (4) y (5) tenemos $f(x^*) = L(x^*,\lambda^*) = MP(0)$.

Concluimos que $f(x^*) = MP(0)$, y como x^* es factible para P, x^* resuelve P.

■

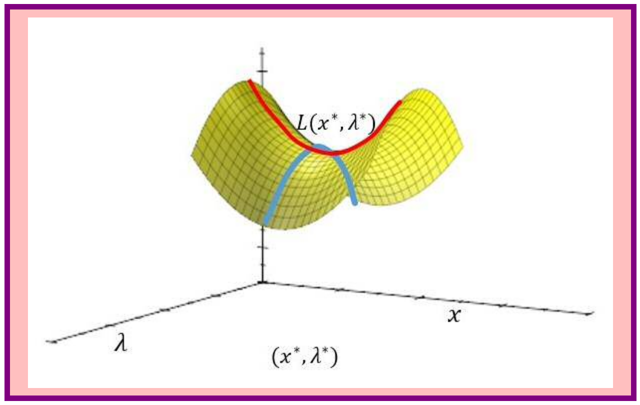

Figura 4.1: Representación gráfica de la función Lagrangeana para las condiciones Teorema de Karush-Kuhn-Tucker en su forma de punto silla.

El la imagen podemos obervar que para cualquier punto de la forma (x^*, λ^*) donde su imagen está dada por $L(x^*, \lambda^*)$ la condición $L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$ pues la parábola formada de color rojo muestra todas la posibles imagenes de x, mientras que la parábola de color azul muestra todos las posibles imagenes de λ.

4.2.2. La forma gradiente del teorema.

***Teorema* 4.5.** *Sean $f : C \to \mathbb{R}$ y $g_i : C \to \mathbb{R}$ con $i = 1, ..., l$, funciones convexas definidas en un subconjunto convexo cerrado no vacío C de $\mathbb{R}^n$. Si definimos a P como el problema de minimizar a $f(x)$ sujeto a la condición $g(x) \leq 0$ con P superconsistente, entonces x^* resuelve P si y sólo si hay un vector λ^* tal que*

1. $\lambda^* \geq 0$,
2. $\lambda_i^* g_i(x^*) = 0$, *para todos* $i = 1, ..., l$,
3. $\bigtriangledown f(x^*) + \sum_{i=1}^{l} \lambda_i^* \bigtriangledown g_i(x^*) = 0$.

Demostración.

Si x^* es una solución para P entonces por el teorema 4.2.1 hay un $\lambda^* \in \mathbb{R}^l$ que satisface (1) y (2) para (x^*, λ^*) en el punto silla de Lagrange en P, entonces

$$L(x^*, \lambda^*) \leq L(x, \lambda^*).$$

Lo anterior se cumple para todo $x \in C$, por lo que x^* es un mínimo global de (x^*, λ^*) en C. Puesto que x^* es un punto interior de C y como $h(x) = L(x, \lambda^*)$ tiene su primera derivada parcial continua en C, se deduce que $\bigtriangledown h(x^*) = 0$, es decir, el gradiente de la condición (3) se cumple.

Por el contrario supongamos que $x^* \in C$ y $\lambda^* \in \mathbb{R}^l$ que satifacen las condiciones (1), (2) y (3). Si x es un punto factible para P. Entonces

$$f(x^*) + \sum_{i=1}^{l} \lambda_i^* g_i(x^*) \leq f(x^*)$$

para $\lambda^* \geq 0$ y $g_i(x^*) \leq 0$ para todo $i = 1, ..., l$. Como f y g_i son funciones continuas y dado que cumplen las condiciones del teorema 1.7, tenemos que para la función f se cumple

$$f(x^*) + \bigtriangledown f(x^*)(x - x^*) \leq f(x) \tag{4}$$

de la misma forma para la función g_i

$$g_i(x^*) + \bigtriangledown g_i(x^*)(x - x^*) \leq g_i(x), \tag{5}$$

multiplicando por λ_i^* y sumando para cada i,

$$\sum_{i=1}^{l} \lambda_i^* g_i(x^*) + \sum_{i=1}^{l} \lambda_i^* \bigtriangledown g_i(x^*)(x - x^*) \leq \sum_{i=1}^{l} \lambda_i^* g_i(x) \tag{6}$$

sumando la ecuación (4) y (6) tenemos

$$\left[f(x^*) + \sum_{i=1}^{l} \lambda_i^* g_i(x^*) \right] + \left[\bigtriangledown f(x^*) + \sum_{i=1}^{l} \lambda_i^* \bigtriangledown g_i(x^*) \right] (x - x^*) \leq f(x) + \sum_{i=1}^{l} \lambda_i^* g_i(x) \leq f(x),$$

como $\lambda_i^* g_i(x^*) = 0$ y $\bigtriangledown f(x^*) + \sum_{i=1}^{l} \lambda_i^* \bigtriangledown g_i(x^*) = 0$, entonces

$$f(x^*) \leq f(x).$$

Así x^* es un mínimo global para $f(x)$, es decir x^* es una solución para P.

■

Ejemplo. *Definamos al problema P como: minimizar $f(x,y) = x^2 - 2x + y^2 + 1$ sujeto a las condiciones $g_1(x,y) = x + y \leq 0$ y $g_2(x,y) = x^2 - 4 \leq 0$.*

Solución.

Notemos que la función $f(x,y)$ y $g(x,y) = (g_1(x,y), g_2(x,y))$ son funciones convexas, para encontrar al punto (x,y) que satisface P debemos encontrar un $\lambda = (\lambda_1, \lambda_2)$ que satisface las condiciones del teorema, es decir, tenemos el siguiente sistema de ecuaciones:

1. $\lambda_1 \geq 0$,
2. $\lambda_2 \geq 0$,
3. $\lambda_1 (x + y) = 0$,
4. $\lambda_2 (x^2 - 4) = 0$,
5. $(2x - 2, 2y) + \lambda_1 (1,1) + \lambda_2 (2x, 0) = (0,0)$.

Desarrollando la ecuación (5) tenemos

$$(2x - 2 + \lambda_1 + 2\lambda_2 x, 2y + \lambda_1) = (0,0),$$

es decir,

$$2x - 2 + \lambda_1 + 2\lambda_2 x = 0 \qquad (6)$$

y

$$2y + \lambda_1 = 0. \qquad (7)$$

De la ecuación (7) despejamos a y, así

$$y = \frac{-\lambda_1}{2},$$

sustituyendo a y en la ecuación (3) y despejando a x se tiene

$$x = \frac{\lambda_1}{2},$$

de igual forma. Sustituyendo a x en la ecuación (4) y despejando a λ_1 obtenemos

$$\lambda_1 = 1,$$

por tanto $x = \frac{1}{2}$ y $y = \frac{-1}{2}$. sustituyendo a x, y y λ_1 en la ecuación (6) y despejando a λ_2 concluimos que

$$\lambda_2 = 0.$$

Por lo tanto el punto $\left(\frac{1}{2}, \frac{-1}{2}, 1, 0\right)$ es punto factible que cumple con las condiciones del teorema.

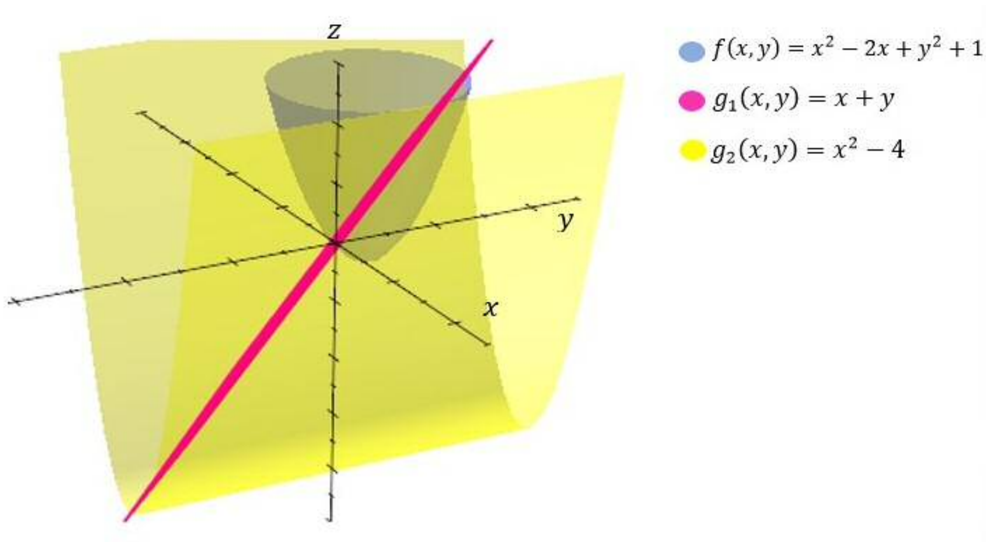

Figura 4.2: Representación gráfica.

Para conocer un poco más podemos consultar [5] y [13].

Bibliografía

[1] Angust Deator and Jhon Muellbauer., Economics and Consumer Behavitor. Cambridge University Press. 1980.

[2] Apostol, T. M., Análisis Matemático, Reverté. México 1960.

[3] Borwein, J. and Lewis, A., Convex Analysis and Nonlinear Optimization. Canadian Mathematical Society Books in Mathematics, New York: Springer-Verlag. 2000.

[4] Censor, Y. and Zenios, S.A., Parallel Optimization: Theory, Algorithms and Applications. New York: Oxford University Press. 1997.

[5] Byrne, C.L., A First Course in Optimization. Department of Mathematical Sciences. University of Massachusetts Lowell. 2014.

[6] Fiacco, A. and McCormick, G., Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Philadelphia, PA: SIAM Classics in Mathematics (reissue). 1990.

[7] Folke Kafka., Teoría Económica. Universidad del Pacífico Centro de Investigación (CIUP). 1981.

[8] Grafe, J., Matemáticas para economistas. McGraw-Hill, Madrid. 1991.

[9] Hancock, H., Theory of maxima and minima. Boston. 1917.

[10] Kalman, D., Leveling with Lagrange: An alternate view of constrained optimization. Mathematics Magazine.2009

[11] McShane, J., The Lagrange multiplier rule. The American Mathematical Monthly, **80**, 1973. 922-925.

[12] Niven, I., Maxima and Minima Without Calculus. Mathematical Association of America. 1981.

[13] Peressini, A., Sullivan, F., and Uhl, J.: The Mathematics of Nonlinear Programming. New York: Springer-Verlag. 1988.

[14] Rangarajan K. Sundaram, A first Course in Optimization Theory. Cambridge, UK. 1996.

[15] Rudin. W., Principios de Análisis Matemático. McGraw-Hill International Editions. 1976.

[16] Sydsaeter, Knut y Hammond, Peter. Matemáticas para el análisis económico. Prentice Hall, 1996.

★ *La vida es bella.* ★