TEMat

El teorema de Karush-Kuhn-Tucker, una generalización del teorema de los multiplicadores de Lagrange, y programación convexa

Fco. Javier Martínez Sánchez
Facultad de Ciencias,
Universidad de Granada
javims282@gmail.com

Resumen: El presente artículo pretende mostrar una generalización del teorema de los multiplicadores de Lagrange, que resuelve problemas de optimización condicionados solo a restricciones de igualdad. El teorema de Karush-Kuhn-Tucker es una extensión suya que resuelve problemas de optimización condicionados tanto a restricciones de igualdad como de desigualdad. En la primera sección del presente texto, se enuncia y comenta el teorema de Lagrange y se incluye un ejemplo de aplicación. En la segunda sección, se enuncia y se demuestra el teorema que extiende al teorema de Lagrange, incluyendo un ejemplo ilustrativo. En la tercera y última sección, se hace una breve introducción a la programación convexa y cóncava y se prueba la condición suficiente en programación convexa y cóncava.

Abstract: This paper expects to show a generalization of the Lagrange multiplier rule, which solves optimization problems with only equality constraints. The Karush-Kuhn-Tucker theorem is an extention of this result in which inequality constraints are also considered. In the first section of this text, we discuss the Lagrange multiplier rule, including one example. In the second one, we prove the Karush-Kuhn-Tucker theorem, including another example. In the third and last one, we make a brief introduction to convex and concave programming and we prove a sufficient condition in convex and concave programming.

Palabras clave: optimización condicionada, programación no lineal, Lagrange, Karush-Kuhn-Tucker, programación convexa, programación cóncava.

MSC2010: 90C30.

Recibido: 28 de octubre de 2018.
Aceptado: 4 de marzo de 2019.

Agradecimientos: A ti, lector.

Referencia: Martínez Sánchez, Fco. Javier. «El teorema de Karush-Kuhn-Tucker, una generalización del teorema de los multiplicadores de Lagrange, y programación convexa». En: *TEMat*, 3 (2019), págs. 33-44. ISSN: 2530-9633. URL: https://temat.es/articulo/2019-p33.

1. Introducción

A lo largo de la historia, la *optimización* ha resultado fructífera a la hora de resolver numerosos problemas de naturaleza variada tanto en matemáticas como en física o economía, entre otras. La optimización es el campo de las matemáticas dedicado a minimizar o maximizar una determinada función, distinguiéndose dos grandes ramas dentro de esta, a saber: *optimización libre* y *optimización condicionada*.

Hoy en día es habitual, en el curso de Cálculo, estudiar que los valores extremos de una función real y derivable f definida en un intervalo abierto $I \subset \mathbb{R}$ se encuentran entre los puntos $x \in I$ tales que $f'(x) = 0$. Y, dado $n \in \mathbb{N}$, los valores extremos de una función real de varias variables y diferenciable definida en un subconjunto abierto $\Omega \subset \mathbb{R}^n$ se encuentran entre los puntos $\boldsymbol{x} \in \Omega$ tales que $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$.

Dados $n \in \mathbb{N}$ un número natural, $\Omega \subset \mathbb{R}^n$ un subconjunto no vacío de $\mathbb{R}^n$ y $f : \Omega \to \mathbb{R}$ una función real definida en Ω, la optimización libre trata de resolver el siguiente problema:

$$\begin{cases} \text{minimizar/maximizar } f(\boldsymbol{x}) \\ \quad \boldsymbol{x} \in \Omega. \end{cases}$$

Pero ¿qué pasa cuando se buscan el mínimo y máximo de una función condicionados a ciertas restricciones o ligaduras? La optimización condicionada se encarga de responder a esta pregunta. En los siglos xvii y xviii, grandes matemáticos (en especial, Lagrange) se ocuparon de obtener máximos y mínimos condicionados de determinadas funciones. A mediados del siglo xviii, Lagrange publicó un método para resolver tales problemas de optimización condicionada solo a restricciones de igualdad: el *método de los multiplicadores de Lagrange*.

El objetivo del presente texto es mostrar una extensión del teorema de Lagrange que sirva para resolver un programa mixto, que es un problema de optimización condicionada donde se minimiza o maximiza una función sujeta a ambos tipos de restricciones: de igualdad y de desigualdad.

La *programación lineal* versa sobre la resolución del programa mixto lineal, esto es, el problema de optimización condicionada a restricciones mixtas (de igualdad y desigualdad)

$$(\star) \qquad \begin{cases} \text{minimizar/maximizar } f(\boldsymbol{x}) \text{ sujeto a} \\ \quad f_1(\boldsymbol{x}) = 0, \dots, f_k(\boldsymbol{x}) = 0, f_{k+1}(\boldsymbol{x}) \leq 0, \dots, f_m(\boldsymbol{x}) \leq 0, \\ \quad \boldsymbol{x} \in \Omega, \end{cases}$$

donde $\Omega \subset \mathbb{R}^n$, $n, k, m \in \mathbb{N}$ y todas las funciones involucradas $f, f_1, \dots, f_m$ son funciones lineales.

Exceptuando al matemático francés G. Monge (1746-1818), quien en 1776 se interesó por problemas de este tipo, debemos remontarnos al año 1939 para encontrar nuevos estudios relacionados con los métodos de la actual programación lineal, entre los que destacan los siguientes matemáticos:

Leonid V. Kantoróvich (1912-1986) publicó una extensa monografía titulada «Mathematical methods of organizing and planning production» [4] en 1939, en la que se hace corresponder una extensa gama de problemas con una teoría matemática concisa.

Tjalling C. Koopmans (1910-1985) formuló el *problema de transporte* para conseguir determinar los planes de embarque al mínimo coste total, conociendo de antemano la disponibilidad y demanda de cada puerto, con la ayuda de Kantarovitch. Ambos fueron galardonados con el Premio Nobel de Economía en 1975.

George B. Dantzig (1914-2015) desarrolló un método iterativo y eficaz de resolución del programa lineal, llamado *método símplex*, utilizado para resolver el problema del puente aéreo de Berlín. Dantzig recibió el Premio de Teoría John von Neumann de la Sociedad Americana de Investigación Operativa del año 1975.

John von Neumann (1903-1957) estableció los fundamentos matemáticos de la programación lineal en 1947, al relacionar esta con la teoría de juegos, que había publicado tres años antes, junto con Oskar Morgenstern, en el libro *Theory of Games and Economic Behavior* [11].

Martínez Sánchez [9] y Dantzig [3] ofrecen una información más ampliada de lo expuesto arriba. Este trabajo trata sobre el teorema fundamental dedicado a resolver el programa mixto general (como el programa ($\star$) pero donde las funciones involucradas no tienen por qué ser lineales).

2. El teorema de Langrange

En esta sección, se define la noción de programa con restricciones de igualdad, se enuncia el teorema de Lagrange (utilizado para resolver tales programas) y se muestra un ejemplo ilustrativo.

Definición 1 (programa con restricciones de igualdad). Dados $n, m \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ un subconjunto abierto y no vacío de $\mathbb{R}^n$ y $m+1$ funciones reales $f, g_1, \dots, g_m$ de clase $\mathcal{C}^1$ definidas en Ω, un **programa con restricciones de igualdad** es un problema de optimización condicionada de la forma

$$\text{(PI)} \qquad \begin{cases} \text{min/max } f(\boldsymbol{x}) \text{ sujeto a} \\ \quad g_1(\boldsymbol{x}) = 0, \dots, g_m(\boldsymbol{x}) = 0, \\ \quad \boldsymbol{x} \in \Omega. \end{cases}$$

Lagrange publicó de manera oficial su resultado en su obra *Mécanique analytique* [8] de 1788 (aunque obtuvo el resultado en agosto 1755 cuando se lo envió por carta a su amigo L. Euler). A continuación se enuncia la versión del teorema de Lagrange del libro de Apostol [1].

Teorema 2 (Lagrange, 1788). *En la situación de la definición 1, si $m < n$, f tiene un extremo condicionado por $g_1(\boldsymbol{x}) = 0, \dots, g_m(\boldsymbol{x}) = 0$ en $\boldsymbol{x}^* \in \Omega$ y la matriz jacobiana de $\mathbf{g} = (g_1, \dots, g_m)$ en $\boldsymbol{x}^*$ tiene rango máximo m, entonces existen m números reales $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ tales que*

$$\nabla f(\boldsymbol{x}^*) + \sum_{k=1}^{m} \lambda_k \nabla g_k(\boldsymbol{x}^*) = \mathbf{0}. \tag{1}$$

La demostración mas común de este teorema hace uso del teorema de la función implícita, hecho por el cual se exige como hipótesis que $m < n$, y puede verse en el libro de Apostol [1]. Sin embargo, esta hipótesis quedará eliminada en la versión general del teorema (teorema de Karush-Kuhn-Tucker). En honor a Lagrange, la ecuación (1) recibe el nombre de **condición de Lagrange** y los escalares $\lambda_1, \dots, \lambda_m$ se denominan **multiplicadores de Lagrange**. Es conveniente advertir que el recíproco del teorema de Lagrange no es cierto: es posible que la condición de Lagrange tenga solución $\boldsymbol{x}^* \in \Omega$ pero que $\boldsymbol{x}^*$ no sea ni mínimo ni máximo de f condicionado a las restricciones de igualdad $g_1(\boldsymbol{x}) = 0, \dots, g_m(\boldsymbol{x}) = 0$. Es conocido que los multiplicadores de Lagrange tienen una interpretación económica. No es el objetivo de este trabajo exponer esta interpretación, pero si el lector lo desea podrá encontrar información al respecto en el libro de Sydsaeter y Hammond [14].

A continuación, se expone un ejemplo sencillo de aplicación del teorema de Lagrange.

Ejemplo 3. Optimización de la función $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x, y) = 2xy$ para todo $(x, y) \in \mathbb{R}^2$ sujeta a la restricción $g(x, y) = x^2 + y^2 \leq 1$.

La función f es continua y el conjunto $K = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ es compacto, luego el teorema de Weierstrass garantiza la existencia de mínimo y máximo de f en K. A la hora de resolver el problema, se distinguen dos casos según si el punto donde f tiene un extremo global pertenece al interior o a la frontera de K.

Por un lado, si el punto donde f tiene un extremo global pertenece al interior de K, entonces se aplica la condición necesaria de existencia de extremo en puntos interiores: las derivadas parciales de f en dicho punto deben anularse, dando lugar a un único punto candidato a extremo:

$$\nabla f(x, y) = 0 \iff \begin{cases} 2y = 0, \\ 2x = 0 \end{cases} \iff (x, y) = (0, 0).$$

Por otro lado, si el punto donde f tiene un extremo global pertenece a la frontera de K, $\mathbb{S} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, entonces se aplica el método de los multiplicadores de Lagrange:

$$\text{Condición de Lagrange: } \begin{cases} 2y + 2\lambda x = 0, \\ 2x + 2\lambda y = 0, \\ x^2 + y^2 = 1 \end{cases} \implies \lambda = \pm 1. \qquad \text{Solución: } \begin{cases} \lambda = 1 \implies \left(\pm\frac{\sqrt{2}}{2}, \mp\frac{\sqrt{2}}{2}\right), \\ \lambda = -1 \implies \left(\pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{2}}{2}\right). \end{cases}$$

Basta comprobar los valores que toma f en cada uno de los puntos obtenidos,

$$f(0,0) = 0, \qquad f\left(\pm\frac{\sqrt{2}}{2}, \mp\frac{\sqrt{2}}{2}\right) = -1, \qquad f\left(\pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{2}}{2}\right) = 1,$$

y se concluye entonces que el mínimo de f condicionado a $x^2 + y^2 \leq 1$ vale -1 y se alcanza en los puntos $\left(\pm\frac{\sqrt{2}}{2}, \mp\frac{\sqrt{2}}{2}\right)$ y que el máximo de f condicionado a $x^2 + y^2 \leq 1$ vale 1 y se alcanza en los puntos $\left(\pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{2}}{2}\right)$. ◂

Las aplicaciones del teorema de Lagrange son muy variadas y numerosas. En particular, resulta verdaderamente útil para demostrar resultados relevantes de análisis matemático como, por ejemplo, las constantes de equivalencia óptimas entre las normas $\|\cdot\|_1$ y $\|\cdot\|_\infty$ de $\mathbb{R}^n$, la desigualdad de Cauchy-Schwarz, el hecho de que toda matriz simétrica real es diagonalizable en $\mathbb{R}$ o la desigualdad de Hadamard, así como para resolver problemas de carácter geométrico, entre otros. En el trabajo de fin de grado de Martínez Sánchez [9], el lector encontrará todo lo anterior. Además, Wu y Wu [15] realizan varias demostraciones de la desigualdad de Cauchy-Schwarz bastante sorprendentes por su sencillez.

Al intentar resolver un problema de extremos condicionados por el método de los multiplicadores de Lagrange, teóricamente es sencillo determinar el sistema de Lagrange asociado, pero en la práctica no siempre existe un procedimiento simple y rápido para resolverlo de manera exacta. En esa situación, una posibilidad es aplicar métodos numéricos con los que se obtengan buenas aproximaciones de la solución del sistema de Lagrange. El lector puede consultar algunos de estos métodos numéricos en el libro de Peressini, Sullivan y Uhl [12].

Ahora bien, ¿qué ocurre cuando se incluyen restricciones de desigualdad en el programa (PI)? El teorema de Karush-Kuhn-Tucker tiene la respuesta, como se verá en la siguiente sección.

3. El teorema de Karush-Kuhn-Tucker

El *teorema de Karush-Kuhn-Tucker* es el primer y principal resultado de toda una teoría que se desarrolló a su alrededor y que dio lugar posteriormente a la *programación no lineal*. En esta sección se enuncia y demuestra dicho teorema, acompañado de un ejemplo ilustrativo.

Antes de enunciar y demostrar el mencionado teorema, veamos sus distintos orígenes históricos, así como su relación con la Segunda Guerra Mundial. En lo que sigue, se hará un breve resumen del artículo de Kjeldsen [6]. Si el lector está interesado puede consultar dicho artículo para profundizar aún más en este tema. Básicamente, el teorema de Karush-Kuhn-Tucker tuvo dos orígenes muy distintos.

En primer lugar, hay que hablar del matemático estadounidense William Karush (1917-1997), quien cursó los estudios de matemáticas en la Universidad de Chicago y cuyo trabajo de fin de máster tenía como título «Minima of functions of several variables with inequalities as side conditions» (1939) [5]. La motivación de Karush era extender un artículo publicado el año anterior por el que, en aquel momento, era el jefe del Departamento de Matemáticas de la Universidad de Chicago, A. Bliss, y que tenía por título «Normality and abnormality in the calculus of variations» [2]. El resultado que demostró Karush en su trabajo en 1939 pertenece indudablemente al campo de la programación no lineal, pero esta área no existía en aquel momento.

El Departamento de Matemáticas de la Universidad de Chicago, fundado con la apertura de la misma en 1892, estaba dirigido por E. Moore junto con G. Bolza y H. Maschke, quienes condujeron al departamento a ser uno de los mas influyentes en matemáticas en Estados Unidos, especialmente en cálculo de variaciones. Bolza estaba profundamente interesado en esta rama de las matemáticas[1] y creó un amplio y fuerte grupo de investigación dedicado única y exclusivamente al cálculo de variaciones. Este grupo fue conocido posteriormente como la *Escuela de Chicago* en cálculo de variaciones y estaba formada tanto por profesores como por alumnos interesados.

[1] Hay que mencionar que Bolza se interesó en el cálculo de variaciones tras asistir a una conferencia de K. Weierstrass en 1879.

En 1908, Maschke falleció y, dos años después, Bolza regresó a Alemania, su país natal. Chicago perdió así a dos líderes matemáticos, lo que se tradujo en un declive que acabó con la llegada al departamento de un nuevo equipo liderado por A. Bliss. Entre 1927 y 1941, el nuevo departamento y, sobre todo, Bliss, que fue alumno de Bolza, continuaron con la tradición de los anteriores líderes y, de nuevo, el departamento se caracterizó por un intenso estudio en cálculo de variaciones que ocupó la mayor parte de la investigación matemática en Chicago. De hecho, entre 1927 y 1937, Bliss dirigió treinta y cinco tesis doctorales, de las cuales treinta y cuatro pertenecían al cálculo de variaciones.

Como estudiante en Chicago, Karush fue producto de esta tradición. En su trabajo, Karush demostró una condición necesaria para la existencia de mínimo local de una función de varias variables $f = f(x_1, \dots, x_n)$ sujeta a desigualdades de la forma $g_1(x) \geq 0, \dots, g_m(x) \geq 0$ con $n, m \in \mathbb{N}$. Karush llevó a cabo su trabajo en 1939 mientras que la Escuela de Chicago se centraba en problemas de cálculo de variaciones con restricciones de desigualdad, donde se minimizaban funcionales de la forma

$$\mathcal{F}[\varphi] = \int_a^b F(x, \varphi(x), \varphi'(x))\, \mathrm{d}x$$

en el conjunto $D = \{\varphi \in \mathcal{C}^1(a, b) : \varphi(a) = A, \varphi(b) = B\}$ para $A, B \in \mathbb{R}$ fijos, $a < b$ y F una función conocida.

Así pues, el trabajo de Karush fue concebido como una versión finitodimensional de los problemas que se atacaban en el cálculo de variaciones y, por lo tanto, en el ambiente de la Escuela de Chicago, carecía de interés y pasó desapercibido. Nadie lo animó a publicarlo y quedó en el olvido durante muchos años.

En segundo lugar, encontramos a los matemáticos Albert W. Tucker (Canadá, 1905-1995) y Harold W. Kuhn (California, 1925-2014), que eran, respectivamente, profesor y alumno en la Universidad de Princeton. Kuhn y Tucker dieron una conferencia en verano de 1950 en Berkeley (Simposio de Berkeley), California, donde enunciaron y demostraron su descubrimiento (lo que hoy en día conocemos por teorema de Karush-Kuhn-Tucker).

En esta conferencia aparece por primera vez en la historia el nombre *programación no lineal*. El objetivo de Kuhn y Tucker era generalizar la programación lineal, que ya había surgido años antes de la mano de Dantzig. A diferencia de Karush, Kuhn y Tucker no tuvieron ningún inconveniente, adquirieron fama casi instantánea en el mundo de las matemáticas y la gente empezó a referirse al resultado como el teorema de Kuhn-Tucker a secas.

Lo que Kuhn y Tucker no sabían es que su resultado no era para nada novedoso. Karush, once años antes, obtuvo prácticamente lo mismo, solo que él utilizaba otras herramientas y notación en la demostración. En 1975, cuando Kuhn y Tucker se enteraron de que el teorema ya había sido probado por Karush en 1939, le escribieron inmediatamente por carta para reconocer su trabajo y prioridad en este asunto.

Se trata, pues, de un descubrimiento múltiple y hoy en día la comunidad matemática se refiere al resultado como **teorema de Karush-Kuhn-Tucker**, apareciendo el apellido Karush en primer lugar ya que este lo demostró once años antes que Kuhn y Tucker; pero, ¿por qué el resultado de Karush pasó desapercibido y tan solo once años después el mismo resultado tuvo tanta fama y reconocimiento? Pues bien, la respuesta a esta pregunta, según la profesora e historiadora de matemáticas T. H. Kjeldsen, está en el contexto histórico y social en el que nació el teorema. El fin de la Segunda Guerra Mundial supuso también la igualdad entre matemática pura y matemática aplicada. Antes de la guerra, la matemática pura era la que gobernaba y dominaba el mundo de las matemáticas pero, durante la guerra, muchos matemáticos dedicados hasta entonces a la investigación en matemática aplicada fueron contratados por diversas organizaciones involucradas en la guerra, como, por ejemplo, el Ejército, para que diseñaran métodos de planificación de programas, una herramienta de las Fuerzas Armadas para llevar a cabo enormes planteamientos logísticos.

Es más, el propio G. Dantzig, contratado por las Fuerzas Armadas en 1941, fue el principal responsable de lo que hoy en día se conoce como programación lineal (que nació durante la guerra bajo el nombre de *programación en estructura lineal*) y del famoso y sencillo método que resuelve un programa lineal, a saber, el *método símplex* (ideado por el propio Dantzig). Más información sobre este método puede verse en el artículo de Peressini, Sullivan y Uhl [12] si el lector está interesado.

Tras la guerra, Dantzig utilizó sus estudios acerca de la programación lineal para resolver el problema del puente aéreo de Berlín: a mediados de 1948, en plena Guerra Fría, la URSS bloqueó las comunicaciones terrestres entre las zonas occidentales alemanas ocupadas por los Aliados y la ciudad de Berlín, y Dantzig, utilizando la programación lineal, diseñó un plan de abastecimiento aéreo minimizando los costes que, en pocos meses, consiguió igualar a los suministros realizados por carretera y ferrocarril antes del bloqueo.

Los países se percataron de que para conseguir ser una potencia mundial debían ser fuertes no solo en el ejército, sino también en ciencia e investigación, y fue ahí donde aumentó la importancia de la matemática aplicada y, en especial, de la programación lineal. Tiene sentido, entonces, que, en ese ambiente, la conferencia sobre programación no lineal de Kuhn y Tucker en 1950 fuera acogida de manera excelente y que, sin embargo, al trabajo de Karush (anterior a la Segunda Guerra Mundial) no se le diera la importancia correspondiente. Kjeldsen [6] y Kuhn [7] ofrecen información sobre todo esto y más. Además, Prékopa [13] permite al lector encontrar la relación de estos problemas de optimización con principios de la física, así como la demostración del *lema de Farkas*, una de las principales herramientas que permitió a Karush demostrar su resultado.

Volviendo ya al contenido matemático, se definen a continuación lo que se entiende por un programa mixto, un punto factible y un punto regular, con lo que estaremos en disposición de enunciar y demostrar el teorema de Karush-Kuhn-Tucker.

Definición 4 (programa mixto). Dados $n, p, q \in \mathbb{N}$, Ω un subconjunto abierto y no vacío de $\mathbb{R}^n$ y $1 + p + q$ funciones reales $f, g_1, \dots, g_p, h_1, \dots, h_q$ de clase $\mathcal{C}^1$ definidas en Ω, un **programa mixto** es un problema de optimización condicionada de la siguiente forma:

$$(\mathrm{PM}^-)\quad \begin{cases} \text{minimizar } f(\boldsymbol{x}) \text{ sujeto a} \\ \quad g_1(\boldsymbol{x}) = 0, \dots, g_p(\boldsymbol{x}) = 0, \\ \quad h_1(\boldsymbol{x}) \le 0, \dots, h_q(\boldsymbol{x}) \le 0, \\ \quad \boldsymbol{x} \in \Omega; \end{cases} \qquad (\mathrm{PM}^+)\quad \begin{cases} \text{maximizar } f(\boldsymbol{x}) \text{ sujeto a} \\ \quad g_1(\boldsymbol{x}) = 0, \dots, g_p(\boldsymbol{x}) = 0, \\ \quad h_1(\boldsymbol{x}) \le 0, \dots, h_q(\boldsymbol{x}) \le 0, \\ \quad \boldsymbol{x} \in \Omega. \end{cases}$$

En lo que sigue, escribiremos (PM) para referirnos indistintamente a (PM^-) o (PM^+). ◂

La función f se denomina *función objetivo* del problema (PM) y las igualdades $\{g_i(\boldsymbol{x}) = 0 : i = 1, \dots, p\}$ y desigualdades $\{h_j(\boldsymbol{x}) \le 0 : j = 1, \dots, q\}$ se denominan *restricciones* de igualdad y desigualdad de (PM), respectivamente.

Definición 5 (punto factible). Un **punto factible** para el problema (PM) es un punto $\boldsymbol{x}^* \in \Omega$ tal que $g_i(\boldsymbol{x}^*) = 0$ y $h_j(\boldsymbol{x}^*) \le 0$ para todo $i = 1, \dots, p$ y $j = 1, \dots, q$. ◂

Definición 6 (punto regular). Un **punto regular** para el problema (PM) es un punto factible $\boldsymbol{x}^* \in \Omega$ de (PM) tal que el conjunto de vectores siguiente es linealmente independiente:

$$\{\nabla g_i(\boldsymbol{x}^*), \nabla h_j(\boldsymbol{x}^*) \in \mathbb{R}^n : i \in \{1, \dots, p\}, j \in J(\boldsymbol{x}^*)\},$$

donde $J(\boldsymbol{x}^*) = \{k \in \{1, \dots, q\} : h_k(\boldsymbol{x}^*) = 0\}$. ◂

Nota 7. Nótese que, bajo las hipótesis del teorema de Lagrange, el punto $\boldsymbol{x}^*$ es regular. ◂

La siguiente versión del teorema de Karush-Kuhn-Tucker y la demostración aquí expuestas no son más que una traducción del artículo de McShane [10].

Teorema 8 (Karush-Kuhn-Tucker, 1939 y 1950). *En la situación de la definición 4, si f tiene un mínimo (resp. máximo) condicionado por $g_i(\boldsymbol{x}) = 0$ para $i = 1, \dots, p$ y por $h_j(\boldsymbol{x}) \le 0$ para $j = 1, \dots, q$, entonces existen números reales $\lambda_0, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q \in \mathbb{R}$, no todos nulos, tales que*

$$\lambda_0 \nabla f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\boldsymbol{x}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\boldsymbol{x}) = \mathbf{0}. \tag{2}$$

Además,

(I) $\lambda_0 \ge 0$ *y* $\mu_j \ge 0$ *(resp.* $\mu_j \le 0$*) para todo* $j = 1, \dots, q$.

(II) $\mu_j h_j(\boldsymbol{x}^*) = 0$ *para* $j = 1, \dots, q$.

(III) *Si* $\boldsymbol{x}^*$ *es un punto regular, entonces se puede tomar* $\lambda_0 = 1$.

Demostración. A continuación, se demuestra el teorema para el caso de mínimo condicionado. El caso de máximo condicionado se deduce de inmediato de este ya que basta aplicarlo a la función $-f$.

Sin pérdida de generalidad asumimos que $\boldsymbol{x}^* = \mathbf{0}$, $f(\boldsymbol{x}^*) = 0$ (esto siempre es posible tras considerar una traslación adecuada) y que, para algún $z \in \mathbb{N}$ con $0 \leq z \leq q$, se tiene $h_1(\boldsymbol{x}^*) = 0, \dots, h_z(\boldsymbol{x}^*) = 0$ y $h_{z+1}(\boldsymbol{x}^*) < 0, \dots, h_q(\boldsymbol{x}^*) < 0$ (basta ordenar las funciones $h_1, \dots, h_q$ de forma que las z primeras se anulen en $\boldsymbol{x}^*$ y las restantes sean menores que cero en $\boldsymbol{x}^*$). Nótese que el caso $z = 0$ corresponde a que en todas las desigualdades se dé la desigualdad estricta, $h_j(\boldsymbol{x}^*) < 0$ para todo $j = 1, \dots, q$, mientras que el caso $z = q$ corresponde a que en todas se dé la igualdad, $h_j(\boldsymbol{x}^*) = 0$ para todo $j = 1, \dots, q$.

Por ser $\boldsymbol{x}^* = \mathbf{0}$ un punto interior de Ω, existe $\varepsilon > 0$ tal que la bola abierta $B(\mathbf{0}, \varepsilon)$ de centro el origen y radio ε está contenida en Ω, y es inmediato que la bola cerrada $\overline{B}(\mathbf{0}, \varepsilon_1)$ de centro el origen y radio $\varepsilon_1 = \varepsilon/2$ está contenida en Ω. Por el teorema de conservación del signo, existe $\varepsilon_2 > 0$ tal que las restricciones h_j con $j = z+1, \dots, q$ son negativas en $\overline{B}(\mathbf{0}, \varepsilon_2)$. Sea $\varepsilon_0 = \text{mín}\{\varepsilon_1, \varepsilon_2\}$; entonces, la bola cerrada $\overline{B}(\mathbf{0}, \varepsilon_0)$ está contenida en Ω y las restricciones h_j para $j = z+1, \dots, q$ son negativas en $\overline{B}(\mathbf{0}, \varepsilon_0)$.

Lema 9. *Para cada $\varepsilon > 0$ con $\varepsilon \leq \varepsilon_0$ existe $N \in \mathbb{N}$ tal que*

$$f(\boldsymbol{x}) + \|\boldsymbol{x}\|^2 + N\left(\sum_{i=1}^{p} g_i(\boldsymbol{x})^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x})^2\right) > 0 \qquad \forall\, \boldsymbol{x} \in \mathbb{S}(\varepsilon), \tag{3}$$

donde $h_j^+(\boldsymbol{x}) = \text{máx}\{h_j(\boldsymbol{x}), 0\}$ para cada $j = 1, \dots, z$ y $\mathbb{S}(\varepsilon) = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| = \varepsilon\}$.

Demostración. Razonemos por reducción al absurdo y supongamos que el enunciado es falso:

> Existe $\tilde{\varepsilon} \in (0, \varepsilon_0]$ de forma que a cada $N \in \mathbb{N}$ le corresponde un punto $\boldsymbol{x}_N \in \mathbb{S}(\tilde{\varepsilon})$ con $f(\boldsymbol{x}_N) + \|\boldsymbol{x}_N\|^2 + N\left(\sum_{i=1}^{p} g_i(\boldsymbol{x}_N)^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x}_N)^2\right) \leq 0$.

Tomamos una sucesión de números naturales $\{N_m\}_{m\in\mathbb{N}}$ creciente y tendiendo a infinito y la sucesión de los correspondientes puntos $\{\boldsymbol{x}_m\}_{m\in\mathbb{N}}$ en $\mathbb{S}(\tilde{\varepsilon})$ de manera que

$$f(\boldsymbol{x}_m) + \|\boldsymbol{x}_m\|^2 + N_m\left(\sum_{i=1}^{p} g_i(\boldsymbol{x}_m)^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x}_m)^2\right) \leq 0 \qquad \forall\, m \in \mathbb{N}. \tag{4}$$

Como $\{\boldsymbol{x}_m\}_{m\in\mathbb{N}}$ es una sucesión acotada de vectores de $\mathbb{R}^n$, el teorema de Bolzano-Weierstrass garantiza la existencia de una sucesión parcial de $\{\boldsymbol{x}_m\}_{m\in\mathbb{N}}$ convergente a un punto $\boldsymbol{x}^0 \in \mathbb{R}^n$. Supongamos, sin pérdida de generalidad, que dicha sucesión parcial es desde un principio la sucesión original $\{\boldsymbol{x}_m\}_{m\in\mathbb{N}}$ y, en virtud de la continuidad de la función objetivo f y de la función norma en $\mathbb{R}^n$, se tiene que

$$\text{lím}\, f(\boldsymbol{x}_m) = f(\boldsymbol{x}^0) \text{ y}$$

$$\|\boldsymbol{x}^0\| = \|\,\text{lím}\, \boldsymbol{x}_m\| = \text{lím}\, \|\boldsymbol{x}_m\| = \text{lím}\, \tilde{\varepsilon} = \tilde{\varepsilon}.$$

Dividiendo ahora ambos miembros de la desigualdad (4) por N_m obtenemos que

$$\frac{f(\boldsymbol{x}_m)}{N_m} + \frac{\|\boldsymbol{x}_m\|^2}{N_m} + \left(\sum_{i=1}^{p} g_i(\boldsymbol{x}_m)^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x}_m)^2\right) \leq 0 \qquad \forall\, m \in \mathbb{N}.$$

Tomando límite cuando m tiende a infinito en la expresión anterior y usando que $\text{lím}\, N_m = \infty$, $\text{lím}\, f(\boldsymbol{x}_m) = f(\boldsymbol{x}^0)$ y $\text{lím}\, \|\boldsymbol{x}_m\|^2 = \tilde{\varepsilon}^2$, lo que implica que $\text{lím}\, \frac{f(\boldsymbol{x}_m)}{N_m} = \text{lím}\, \frac{\|\boldsymbol{x}_m\|^2}{N_m} = 0$, se sigue que

$$\sum_{i=1}^{p} g_i(\boldsymbol{x}^0)^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x}^0)^2 \leq 0,$$

de donde se deduce que $\boldsymbol{x}^0$ satisface que $g_i(\boldsymbol{x}^0) = 0$ para $i = 1, \dots, p$ y $h_j^+(\boldsymbol{x}^0) = 0$ para $j = 1, \dots, z$. Así que $\text{lím}\, f(\boldsymbol{x}_m) = f(\boldsymbol{x}^0) \geq f(\boldsymbol{x}^* = \mathbf{0}) = 0$, pero por (4) se desprende que $f(\boldsymbol{x}_m) \leq -\tilde{\varepsilon}^2 < 0$ para todo $m \in \mathbb{N}$, luego $\text{lím}\, f(\boldsymbol{x}_m) = f(\boldsymbol{x}^0) < 0$ y se llega a una contradicción. ■

Lema 10. *Para cada* $\varepsilon > 0$ *con* $\varepsilon \leq \varepsilon_0$ *existen un punto* $\overline{\boldsymbol{x}} = (\overline{x}_1, \dots, \overline{x}_n) \in \mathbb{R}^n$ *y un vector unitario* $(\lambda_0, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_z)$ *con componentes* $\lambda_0, \mu_1, \dots, \mu_z$ *no negativas tales que* $\|\overline{\boldsymbol{x}}\| < \varepsilon$ *y*

$$\lambda_0 \left[D_k f(\overline{\boldsymbol{x}}) + 2\overline{x}_k\right] + \sum_{i=1}^{p} \lambda_i D_k g_i(\overline{\boldsymbol{x}}) + \sum_{j=1}^{z} \mu_j D_k h_j(\overline{\boldsymbol{x}}) = 0 \qquad \forall\, k = 1, \dots, n. \tag{5}$$

Demostración. Sea $\tilde{\varepsilon} \in (0, \varepsilon_0]$ fijo pero arbitrario y $N \in \mathbb{N}$ el número natural dado por el lema 9. Considérese la función $F \colon \overline{B}(\mathbf{0}, \tilde{\varepsilon}) \subset \Omega \to \mathbb{R}$ definida por

$$F(\boldsymbol{x}) = f(\boldsymbol{x}) + \|\boldsymbol{x}\|^2 + N\left(\sum_{i=1}^{p} g_i(\boldsymbol{x})^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x})^2\right) \qquad \forall\, \boldsymbol{x} \in \overline{B}(\mathbf{0}, \tilde{\varepsilon}).$$

Como F es continua y $\overline{B}(\mathbf{0}, \tilde{\varepsilon})$, compacto, el teorema de Weierstrass asegura la existencia de un punto $\overline{\boldsymbol{x}} \in \overline{B}(\mathbf{0}, \tilde{\varepsilon})$ donde F alcanza su mínimo global; en particular, $F(\overline{\boldsymbol{x}}) \leq F(\boldsymbol{x}^* = \mathbf{0}) = 0$, y el lema 9 afirma que $\|\overline{\boldsymbol{x}}\| < \tilde{\varepsilon}$ ($\overline{\boldsymbol{x}}$ es un punto interior de $B(\mathbf{0}, \tilde{\varepsilon})$). Así, todas las derivadas parciales de primer orden de F deben anularse en $\overline{\boldsymbol{x}}$:

$$D_k f(\overline{\boldsymbol{x}}) + 2\overline{x}_k + 2N\left(\sum_{i=1}^{p} g_i(\overline{\boldsymbol{x}}) D_k g_i(\overline{\boldsymbol{x}}) + \sum_{j=1}^{z} h_j^+(\overline{\boldsymbol{x}}) D_k h_j(\overline{\boldsymbol{x}})\right) = 0 \qquad \forall\, k = 1, \dots, n, \tag{6}$$

donde se ha usado que la función $(h_j^+)^2$ $(j = 1, \dots, z)$ es diferenciable en $B(\mathbf{0}, \tilde{\varepsilon})$ con

$$D_k (h_j^+)^2(\boldsymbol{x}) = 2h_j^+(\boldsymbol{x}) D_k h_j(\boldsymbol{x}) \qquad \forall\, \boldsymbol{x} \in B(\mathbf{0}, \tilde{\varepsilon})$$

para cualquier $k = 1, \dots, n$, lo cual se demuestra en el libro de Peressini, Sullivan y Uhl [12, capítulo 6] teniendo en cuenta que

$$h_j^+(\boldsymbol{x}) = \frac{h_j(\boldsymbol{x}) + |h_j(\boldsymbol{x})|}{2} \qquad \forall\, \boldsymbol{x} \in B(\mathbf{0}, \tilde{\varepsilon}).$$

Tomando $\tau = \left[1 + 4N^2 \left(\sum_{i=1}^{p} g_i(\overline{\boldsymbol{x}})^2 + \sum_{j=1}^{z} h_j^+(\overline{\boldsymbol{x}})^2\right)\right]^{1/2} > 0$ y definiendo

$$\lambda_0 = \frac{1}{\tau}, \qquad \lambda_i = 2N\frac{g_i(\overline{\boldsymbol{x}})}{\tau} \quad (i = 1, \dots, p), \qquad \mu_j = \begin{cases} 2N\dfrac{h_j^+(\overline{\boldsymbol{x}})}{\tau} & j = 1, \dots, z; \\ 0 & j = z+1, \dots, q, \end{cases}$$

es fácil comprobar que el vector $(\lambda_0, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_z)$ es unitario, λ_0 y μ_j $(j = 1, \dots, z)$ son no negativos y dividiendo ambos miembros de la condición (6) por τ se consigue la igualdad (5). ∎

Finalmente, tomamos una sucesión decreciente de números reales y positivos $\{\delta_m\}_{m \in \mathbb{N}}$ tendiendo a cero con $\delta_1 < \varepsilon_0$. Para cada $m \in \mathbb{N}$, elegimos un punto $\overline{\boldsymbol{x}}_m \in \mathbb{R}^n$ con $\|\overline{\boldsymbol{x}}_m\| < \delta_m$ y un vector unitario $(\lambda_{0,m}, \lambda_{1,m}, \dots, \lambda_{p,m}, \mu_{1,m}, \dots, \mu_{z,m}, 0, \dots, 0)$ con componentes $\lambda_{0,m}$ y $\mu_{j,m}$ $(j = 1, \dots, z)$ no negativas tales que se cumple la igualdad (5) (esto es posible por el lema 10). De nuevo, por el teorema de Bolzano-Weierstrass, existe una sucesión parcial para la cual los vectores unitarios convergen a un límite $(\lambda_0, \lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q)$.

Como $\{\overline{\boldsymbol{x}}_m\}_{m \in \mathbb{N}} \to \boldsymbol{x}^* = \mathbf{0}$, la ecuación (5) se cumple para este vector límite y esto demuestra el teorema salvo el apartado (III), que se comenta a continuación: si $\boldsymbol{x}^*$ es un punto regular, no puede ser que $\lambda_0 = 0$, pues entonces la condición (2) contradice la independencia lineal de los vectores $\nabla g_1(\boldsymbol{x}^0), \dots, \nabla g_p(\boldsymbol{x}^*)$, $\nabla h_1(\boldsymbol{x}^*), \dots, \nabla h_z(\boldsymbol{x}^*)$, y esto concluye la demostración del teorema. ∎

Esta demostración se le debe a Edward J. McShane (1904-1989) y fue publicada en un artículo para *The American Mathematical Monthly* en el año 1973 [10], y es realmente asombrosa: únicamente aplica teoremas elementales del análisis matemático como son el teorema de Weierstrass, el teorema de Bolzano-Weierstrass y la condición necesaria de existencia de extremo en un punto interior.

Nótese que si $q = 0$ y $p > 0$, entonces el problema (PM) incluye solo restricciones de igualdad y el clásico método de los multiplicadores de Lagrange aporta una condición necesaria de existencia de solución del programa, y el caso extremo $p = q = 0$ da lugar a un problema de optimización libre.

En honor a Karush, Kuhn y Tucker, la ecuación (2) junto con los apartados (I) y (II) del teorema 8 reciben el nombre de **condiciones de Karush-Kuhn-Tucker**, los escalares $\lambda_1, \dots, \lambda_p$ (asociados a las restricciones de igualdad) se denominan **multiplicadores de Lagrange** y los escalares $\mu_1, \dots, \mu_q$ (asociados a las restricciones de desigualdad) se denominan **multiplicadores de Karush-Kuhn-Tucker**. Al igual que el teorema de Lagrange, es conveniente advertir que el teorema de Karush-Kuhn-Tucker solo aporta una condición necesaria, y no suficiente, de existencia de solución del programa (PM).

Obsérvese que no se impone ninguna hipótesis sobre el número de variables n y el número de restricciones p y q, a diferencia del teorema de Lagrange, donde se exigía que el número de restricciones de igualdad m fuese menor que el número de variables n. En el caso $q = 0$ (solamente restricciones de igualdad están presentes), las condiciones (I) y (II) del ecuación (2) carecen de sentido.

En la práctica, si se sabe de antemano que existe un punto $\boldsymbol{x}^* \in \Omega$ que es solución de (PM), entonces $\boldsymbol{x}^*$ cumple las siguientes condiciones:

Condición estacionaria: $\lambda_0 \nabla f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\boldsymbol{x}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\boldsymbol{x}) = \boldsymbol{0}$.

Condición de factibilidad: $g_i(\boldsymbol{x}^*) = 0$ para $i = 1, \dots, p$ y $h_j(\boldsymbol{x}^*) \le 0$ para $j = 1, \dots, q$.

Condición de holgura: $\mu_j h_j(\boldsymbol{x}^*) = 0$ para $j = 1, \dots, q$.

Condición de signo: $\lambda_0 \ge 0$ y $\mu_j \ge 0$ para todo $j = 1, \dots, q$ (para mínimo condicionado) o $\mu_j \le 0$ para todo $j = 1, \dots, q$ (para máximo condicionado).

A continuación, se expone un primer ejemplo sencillo de aplicación del teorema de Karush-Kuhn-Tucker.

Ejemplo 11. Resolver el siguiente problema de optimización:

$$\begin{cases} \text{optimizar } f(x, y, z) = x + y + z \text{ sujeto a} \\ \quad h_1(x, y, z) = (y-1)^2 + z^2 \le 1, \\ \quad h_2(x, y, z) = x^2 + (y-1)^2 + z^2 \le 3. \end{cases}$$

Por el teorema de Weierstrass, existe solución (máximo y mínimo) del programa considerado y las condiciones de Karush-Kuhn-Tucker son las siguientes:

Condición estacionaria:

$$\begin{cases} 1 + 2\mu_2 x = 0, \\ 1 + 2\mu_1(y-1) + 2\mu_2(y-1) = 0, \\ 1 + 2\mu_1 z + 2\mu_2 z = 0. \end{cases}$$

Condición de factibilidad:

$$\begin{cases} (y-1)^2 + z^2 \le 1, \\ x^2 + (y-1)^2 + z^2 \le 3. \end{cases}$$

Condición de holgura:

$$\begin{cases} \mu_1 \left[(y-1)^2 + z^2 - 1 \right] = 0, \\ \mu_2 \left[x^2 + (y-1)^2 + z^2 - 3 \right] = 0. \end{cases}$$

Condición de signo:

$$\begin{cases} \mu_1, \mu_2 \ge 0 \longrightarrow \text{mínimo local}, \\ \mu_1, \mu_2 \le 0 \longrightarrow \text{máximo local}. \end{cases}$$

A partir de la condición de holgura se distinguen cuatro casos:

$$\begin{cases} \mu_1 = 0 & \Longrightarrow \begin{cases} \mu_2 = 0, & \text{(caso I)} \\ x^2 + (y-1)^2 + z^2 - 3 = 0; & \text{(caso II)} \end{cases} \\ (y-1)^2 + z^2 - 1 = 0 & \Longrightarrow \begin{cases} \mu_2 = 0, & \text{(caso III)} \\ x^2 + (y-1)^2 + z^2 - 3 = 0, & \text{(caso IV)} \end{cases} \end{cases}$$

pero la primera ecuación de la condición estacionaria obliga a que $\mu_2 \neq 0$, así que los solo se deben comprobar los casos II y IV.

Caso II: para el caso II, un sencillo cálculo más o menos breve proporciona los siguientes dos puntos de Karush-Kuhn-Tucker con sus correspondientes multiplicadores:

$$P_1 = (1, 2, 1), \qquad \boldsymbol{\mu} = \left(0, -\frac{1}{2}\right),$$

$$P_2 = (-1, 0, -1), \qquad \boldsymbol{\mu} = \left(0, \ \frac{1}{2}\right).$$

Caso IV: para el caso IV, un sencillo cálculo más o menos breve proporciona los siguientes cuatro puntos de Karush-Kuhn-Tucker con sus correspondientes multiplicadores:

$$
\begin{aligned}
P_3 &= \left(\sqrt{2}, 1+\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), & \mu &= \left(-\frac{1}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right),\\
P_4 &= \left(\sqrt{2}, 1-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), & \mu &= \left(\frac{3}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right),\\
P_5 &= \left(-\sqrt{2}, 1+\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), & \mu &= \left(-\frac{3}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right),\\
P_6 &= \left(-\sqrt{2}, 1-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), & \mu &= \left(\frac{1}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right).
\end{aligned}
$$

Finalmente, se aplican las dos condiciones aún no usadas (factibilidad y signo) y se resumen los resultados obtenidos en el cuadro 1. ◂

Cuadro 1: Resumen de los resultados del ejemplo 11.

P	μ	Factibilidad	Signo	Conclusión
P_1	$\left(0, -\frac{1}{2}\right)$	NO	-	-
P_2	$\left(0, \frac{1}{2}\right)$	NO	-	-
P_3	$\left(-\frac{1}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right)$	SÍ	Negativo	Máximo condicionado
P_4	$\left(\frac{3}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right)$	SÍ	NO	-
P_5	$\left(-\frac{3}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right)$	SÍ	NO	-
P_6	$\left(\frac{1}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right)$	SÍ	Positivo	Mínimo condicionado

4. Programación convexa y cóncava

En esta última sección se expone lo que se entiende por programa convexo y cóncavo y se prueba que la condición necesaria de existencia de solución para programas mixtos que aporta el teorema de Karush-Kuhn-Tucker es también condición suficiente bajo hipótesis de convexidad o concavidad. Antes de ello, recordemos el siguiente resultado que caracteriza a las funciones diferenciables que son convexas o cóncavas.

Proposición 12 (caracterización de funciones convexas y cóncavas). *Sean $n \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ un subconjunto abierto, convexo y no vacío de $\mathbb{R}^n$ y $f\colon \Omega \to \mathbb{R}$ una función real y diferenciable en Ω. Entonces,*

- *f es convexa si y solo si $f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle \leq f(\boldsymbol{y})$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.*
- *f es estrictamente convexa si y solo si $f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle < f(\boldsymbol{y})$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ con $\boldsymbol{x} \neq \boldsymbol{y}$.*
- *f es cóncava si y solo si $f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle \geq f(\boldsymbol{y})$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.*
- *f es estrictamente cóncava si y solo si $f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle > f(\boldsymbol{y})$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ con $\boldsymbol{x} \neq \boldsymbol{y}$.*

Definición 13 (programa convexo). Dados $n, m \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ un subconjunto abierto, convexo y no vacío de $\mathbb{R}^n$ y $f, h_1, \dots, h_m$ funciones reales, diferenciables y convexas en Ω, un **programa convexo** es un problema de minimización condicionada de la siguiente forma:

$$(\mathrm{PC}^-) \qquad \begin{cases} \text{minimizar } f(\boldsymbol{x}) \text{ sujeto a} \\ \quad h_1(\boldsymbol{x}) \le 0, \dots, h_m(\boldsymbol{x}) \le 0, \\ \quad \boldsymbol{x} \in \Omega. \end{cases}$$ ◂

Definición 14 (programa cóncavo). Dados $n, m \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ un subconjunto abierto, convexo y no vacío de $\mathbb{R}^n$ y $f, h_1, \dots, h_m$ funciones reales, diferenciables y cóncavas en Ω, un **programa cóncavo** es un problema de maximización condicionada de la siguiente forma:

$$(\mathrm{PC}^+) \qquad \begin{cases} \text{maximizar } f(\boldsymbol{x}) \text{ sujeto a} \\ \quad h_1(\boldsymbol{x}) \le 0, \dots, h_m(\boldsymbol{x}) \le 0, \\ \quad \boldsymbol{x} \in \Omega. \end{cases}$$ ◂

Teorema 15 (condición suficiente para programas convexos). *Si $\boldsymbol{x}^* \in \Omega$ es un punto factible y regular para el programa convexo y cumple las condiciones de Karush-Kuhn-Tucker, entonces $\boldsymbol{x}^*$ es solución del programa convexo, esto es, $\boldsymbol{x}^*$ es un mínimo global de f condicionado a $h_k(\boldsymbol{x}) \le 0$ para $k = 1, \dots, m$.*

Demostración. Como $\boldsymbol{x}^*$ es un punto regular por hipótesis, se puede tomar $\lambda_0 = 1$. Por hipótesis, existen escalares $\mu_1, \dots, \mu_m \in \mathbb{R}_0^+$ no todos nulos tales que $\nabla f(\boldsymbol{x}^*) + \sum_{k=1}^m \mu_k \nabla h_k(\boldsymbol{x}^*) = 0$. Como $\mu_k \ge 0$ para cada $k = 1, \dots, m$ y $h_k(\boldsymbol{x}) \le 0$ para cada $\boldsymbol{x} \in \Omega$ y $k = 1, \dots, m$, se tiene que $\mu_k h_k(\boldsymbol{x}) \le 0$ para cada $\boldsymbol{x} \in \Omega$ y $k = 1, \dots, m$, y esto, junto con la proposición 12, permite escribir lo siguiente para cualquier punto factible $\boldsymbol{x} \in \Omega$:

$$\begin{aligned} f(\boldsymbol{x}) &\ge f(\boldsymbol{x}) + \sum_{k=1}^m \mu_k h_k(\boldsymbol{x}) \\ &\ge (f(\boldsymbol{x}^*) + \langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle) + \sum_{k=1}^m \mu_k \left(h_k(\boldsymbol{x}^*) + \langle \nabla h_k(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \right) \\ &= f(\boldsymbol{x}^*) + \langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle + \sum_{k=1}^m \underbrace{\mu_k h_k(\boldsymbol{x}^*)}_{0} + \langle \sum_{k=1}^m \mu_k \nabla h_k(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \\ &= f(\boldsymbol{x}^*) + \langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle + \langle \sum_{k=1}^m \mu_k \nabla h_k(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \\ &= f(\boldsymbol{x}^*) + \langle \underbrace{\nabla f(\boldsymbol{x}^*) + \sum_{k=1}^m \mu_k \nabla h_k(\boldsymbol{x}^*)}_{\boldsymbol{0}}, \boldsymbol{x} - \boldsymbol{x}^* \rangle = f(\boldsymbol{x}^*), \end{aligned}$$

lo que significa que $\boldsymbol{x}^*$ es un mínimo global de f condicionado a $h_k(\boldsymbol{x}) \le 0$ para cada $k = 1, \dots, m$, como se quería. ■

Una demostración análoga a la recién expuesta probaría la condición suficiente para programas cóncavos. Animamos al lector a intentarlo.

Nota 16. En la situación del teorema 15, de la proposición 12 se deduce que, si la función objetivo f de (PC^-) (resp. (PC^+)) es estrictamente convexa (resp. estrictamente cóncava), entonces el mínimo global (resp. máximo global) de (PC^-) (resp. (PC^+)) es único. ◂

Como comentario final, hay que destacar el alcance que tuvo y tienen los resultados aquí expuestos, con mención especial a la programación convexa. La convexidad es una herramienta eficaz que aporta una condición suficiente a la hora de resolver ciertos problemas de optimización y, además, grandes ramas dentro de la programación, como la programación lineal (donde la función a optimizar es una función lineal), geométrica (donde la función a optimizar es un posinomio[2]) o cuadrática (donde la función a

[2] Un posinomio tiene la misma expresión que un polinomio en varias variables $x_1, \dots, x_n$, pero aquí las variables son solo positivas, los coeficientes son solo positivos y los exponentes son reales (positivos, negativos o cero).

optimizar es una función cuadrática), pueden atacarse aplicando el teorema 15. Todo esto puede verse detalladamente en el trabajo final de grado de Martínez Sánchez [9], así como la *programación convexa dual*, que a grandes rasgos, trata de asociar a cada programa convexo de minimización (PC^-) un programa de maximización libre (sin restricciones) llamado *programa convexo dual* y que, a menudo, es mas fácil de resolver que el propio (PC^-) y cuyas soluciones pueden usarse para generar soluciones de (PC^-). Véase el libro de Peressini, Sullivan y Uhl [12] para más información al respecto.

Referencias

[1] Apostol, Tom M. *Análisis matemático*. Trad. por Vélez Cantarell, Francisco. Barcelona: Reverté, 1960.

[2] Bliss, G. A. «Normality and abnormality in the calculus of variations». En: *Transactions of the American Mathematical Society* 43.3 (1938), págs. 365-376. issn: 0002-9947. https://doi.org/10.2307/1990066.

[3] Dantzig, George B. *Linear programming and extensions*. RAND Corporation, 1963, págs. xvi+625. https://doi.org/10.7249/R366.

[4] Kantorovich, L. V. «Mathematical methods of organizing and planning production». En: *Management Science. Journal of the Institute of Management Science. Application and Theory Series* 6 (1959/1960), págs. 366-422. issn: 0025-1909. https://doi.org/10.1287/mnsc.6.4.366.

[5] Karush, William. *Minima of functions of several variables with inequalities as side conditions*. Thesis (SM). The University of Chicago, 1939, pág. 25. url: http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqdiss&rft_dat=xri:pqdiss:TM15121.

[6] Kjeldsen, Tinne Hoff. «A contextualized historical analysis of the Kuhn-Tucker theorem in nonlinear programming: the impact of World War II». En: *Historia Mathematica* 27.4 (2000), págs. 331-361. issn: 0315-0860. https://doi.org/10.1006/hmat.2000.2289.

[7] Kuhn, Harold W. «Nonlinear programming: a historical view». En: *Nonlinear programming Proceedings of a Symposium in Applied Mathematics Held in New York City*. Vol. 9. SIAM-AMS Proceedings. American Mathematical Society, 1976, págs. 1-26.

[8] Lagrange, Joseph-Louis. *Mécanique analytique*. Edición revisada. Librairie Scientifique et Technique Albert Blanchard, 1965.

[9] Martínez Sánchez, Francisco Javier. *Una generalización del teorema de los multiplicadores de Lagrange: condiciones de Karush-Kuhn-Tucker en programación no lineal*. Trabajo de Fin de Grado. Universidad de Granada, 2018. url: https://www.ugr.es/~acanada/docencia/matematicas/TFG-definitivo-2julio2018.pdf.

[10] McShane, Edward J. «The Lagrange Multiplier Rule». En: *The American Mathematical Monthly* 80.8 (1973), págs. 922-925. https://doi.org/10.1080/00029890.1973.11993409.

[11] Neumann, John von y Morgenstern, Oskar. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey, 1944, págs. xviii+625.

[12] Peressini, Anthony L.; Sullivan, Francis E., y Uhl, J. J., Jr. *The mathematics of nonlinear programming*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1988, págs. x+273. isbn: 978-0-387-96614-4.

[13] Prékopa, András. «On the development of optimization theory». En: *The American Mathematical Monthly* 87.7 (1980), págs. 527-542. issn: 0002-9890. https://doi.org/10.2307/2321417.

[14] Sydsaeter, Knut y Hammond, Peter. *Matemáticas para el análisis económico*. Prentice Hall, 1996. isbn: 978-0-13-240615-4.

[15] Wu, Hui-Hua y Wu, Shanhe. «Various proofs of the Cauchy-Schwarz inequality». En: *Octogon Mathematical Magazine* 17.1 (2009), págs. 221-229. url: http://www.uni-miskolc.hu/~matsefi/Octogon/volumes/volume1/article1_19.pdf.