Universidad de Granada

FACULTAD DE CIENCIAS
GRADO EN MATEMÁTICAS

Una generalización del Teorema de los multiplicadores de Lagrange: condiciones de Karush-Kuhn-Tucker en Programación no lineal

Fco. Javier Martínez Sánchez

Departamento de Análisis Matemático
Tutor: Antonio Cañada Villar

Curso 2017/2018

"*Puesto que el Universo es perfecto*
y fue creado por el Creador más sabio,
nada ocurre en él sin que esté presente
alguna ley de máximo o mínimo"

– L. Euler (s. XVIII)

Agradecimientos

Quisiera agradecer este trabajo a mi familia por todo el apoyo recibido durante el transcurso del grado, a mi profesor de Bachillerato, Juan Castilla, que fue el culpable de mi interés hacia las matemáticas y a los profesores Jerónimo Alaminos, Rafael Payá, Armando Reyes Villena, Juan Soler y Antonio Cañada, entre otros. Todos ellos han sido los responsables de mi "atracción" por una de las ramas mas importantes y bellas de las matemáticas, a saber, el Análisis matemático. Por último, me gustaría hacer una mención especial a mi tutor de TFG, Antonio Cañada, del cual he aprendido mucho, he tenido la suerte de disfrutarlo como profesor en mas de una asignatura del grado y siempre ha estado a disposición cuando lo he necesitado para escribir el presente trabajo.

Índice general

Introducción **2**

Resumen **6**

Summary **8**

1. Conocimientos previos sobre extremos libres y los multiplicadores de Lagrange **11**
1.1. Extremos libres para funciones de una variable 11
1.1.1. Extremos locales libres 12
1.1.2. Extremos globales libres 14
1.2. Extremos libres para funciones de varias variables 15
1.2.1. Extremos locales libres 16
1.2.2. Extremos globales libres 17
1.3. El método de los multiplicadores de Lagrange 19

2. Orígenes históricos de la Programación no lineal **33**
2.1. Introducción 33
2.1.1. Prerrequisitos matemáticos 34
2.2. El Teorema de Karush 35
2.3. El Teorema de F. John 39
2.4. El Teorema de Kuhn y Tucker 43
2.5. El aspecto de un descubrimiento múltiple 46
2.6. La importancia del contexto histórico 47
2.7. Notas finales 51

3. El Teorema de Karush-Kuhn-Tucker y aplicaciones en Programación convexa **53**
3.1. Programas con restricciones mixtas 53
3.1.1. La demostración de McShane y ejemplos 55
3.2. Convexidad 63
3.2.1. Conjuntos convexos 63
3.2.2. Funciones convexas 65
3.3. Programación convexa 71
3.4. Programación convexa dual 79
3.4.1. Programación lineal 82

3.4.2. Programación geométrica 84
3.4.3. Programación cuadrática 92

Bibliografía **95**

Introducción

El concepto de derivada presupone los de función y límite funcional, los cuales tuvieron una larga evolución hasta alcanzar su significado actual, por eso la definición de derivada es relativamente reciente. Aún así, técnicas en las que podemos reconocer el uso de derivadas se han venido usando desde el siglo XVII, incluso antes de que Newton y Leibnitz, en el último tercio de dicho siglo, las formularan en términos de *fluxiones* y de cocientes diferenciales, respectivamente. Durante los siglos XVIII y XIX las derivadas fueron ampliamente desarrolladas y aplicadas a campos muy diversos, pero no fueron descritas rigurosamente hasta el último tercio del siglo XIX. Todo este proceso lo resume la historiadora de las matemáticas Judith V. Grabiner en una adecuada cita: *Primero, la derivada fue usada, después descubierta, explorada y desarrollada y, finalmente, definida.* La derivada goza de numerosas aplicaciones prácticas como, por ejemplo, el cálculo de tangentes y de valores extremos para funciones derivables, entre otros (véase [34]). El matemático francés Pierre de Fermat (1601-1665) escribió una memoria titulada *Methodus ad disquirendam maximan et minimam* (Método para la investigación de máximos y mínimos), en la que se establecía el primer procedimiento general para calcular máximos y mínimos de funciones e ilustró su método hallando el punto C del segmento de recta AB que hace máxima el área del rectángulo de lados AC y CB, como sigue:

1. Sean x y $L - x$ los dos segmentos, donde L es la longitud del segmento AB.
2. La función a maximizar es $f(x) = (L - x)x = Lx - x^2$.
3. Se sustituye x por $x + h$ en la expresión de la función f, obteniendo una nueva expresión que depende de x y h, a saber,
$$f_h(x) = (L - x - h)(x + h) = L(x + h) - 2xh - (x^2 + h^2)$$
4. Se igualan ambas expresiones y se simplifica el resultado:
$$Lx - x^2 = L(x + h) - 2xh - (x^2 + h^2) \Longrightarrow Lh = 2xh + h^2$$
5. Se divide por h: $L = 2x + h^2$.
6. Se elimina el término h: $L = 2x$.

Así, la solución obtenida es $C = L/2$ (la mitad del segmento AB).

Es tentador reproducir este razonamiento en términos actuales. Esencialmente, lo que Fermat hacía en los pasos 1-4 era construir la función área del rectángulo obtenido, $f(x) = x(L - x)$ y hacer $f(x + \Delta x) - f(x) \sim 0$. En el paso 5, hacía $\frac{f(x+\Delta x)-f(x)}{\Delta x} \sim 0$ y por último, en el paso 6, $\left(\frac{f(x+\Delta x)-f(x)}{\Delta x}\right)_{\Delta x=0} = 0$. Claro que Fermat no decía nada acerca de que h fuese un infinitesimal, ni siquiera una magnitud muy pequeña, y el método no implica ningún concepto de límite, sino que es puramente algebraico y algorítmico.

Hoy en día, cualquier estudiante de Cálculo sabe que los valores extremos de una función derivable f definida en un intervalo abierto $I \subset \mathbb{R}$ se encuentran entre los puntos

$x \in I$ que cumplen $f'(x) = 0$ y esto es basicamente lo que hacía Fermat en su método (obviamente, sin saber lo que era la derivada de una función, es más, ni siquiera tenía claro el concepto de límite) y únicamente para funciones polinómicas. Aún así, en muchos textos matemáticos de la actualidad a este resultado elemental se le conoce como *lema de Fermat* (teorema 1, Cap. 1). El lema de Fermat constituye el primer gran resultado acerca de optimización de funciones de una variable, el cual se extiende para funciones de varias variables (teorema 5, Cap. 1) y que trata de una condición necesaria (y no suficiente) para la existencia de máximos y mínimos. Esta condición necesaria junto con la condición suficiente (teorema 3 para funciones de una variable y teorema 7 para funciones de varias variables), que suponen mas regularidad a la función en concreto, constituyen las principales herramientas para atacar un problema de *optimización libre*. Dado $n \in \mathbb{N}$ y una función $f : A \subset \mathbb{R}^n \to \mathbb{R}$, un problema de optimización libre consiste en calcular, si existen, los puntos $\boldsymbol{x_0}, \boldsymbol{x^0} \in A$ tales que $f(\boldsymbol{x_0}) \leq f(\boldsymbol{x}) \leq f(\boldsymbol{x^0})$ para todo $\boldsymbol{x} \in A$. La función f que se optimiza se llama *función objetivo* y los puntos $\boldsymbol{x_0}$ y $\boldsymbol{x^0}$ son, respectivamente, los valores mínimo y máximo que alcanza f en A, pero ¿qué pasa cuando se buscan el mínimo y máximo de una función sujetos o condicionados a ciertas ligaduras o restricciones? Este nuevo problema que surge recibe el nombre de problema de *optimización condicionada*. En los siglos XVII y XVIII, grandes matemáticos, en especial J. L. 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*, que se comenta en el primer capítulo de este trabajo. Al igual que el lema de Fermat para optimización libre, este método aporta sólo una condición necesaria para la existencia de mínimos y máximos condicionados a restricciones de igualdad (teorema 8, Cap. 1).

El siguiente paso parece claro: extender el resultado debido a Lagrange al caso en el que se optimiza una función bajo restricciones de igualdad pero también de desigualdad. En un primer acercamiento a este problema, surgió la *Programación lineal*. Aunque la Programación lineal surgió para dar respuesta a problemas logísticos y militares, a lo largo de la historia ha tenido aplicaciones importantes en distintas áreas de la actividad humana, desde el social y estratégico hasta la industria y la economía. El nombre de Programación lineal no procede de la creación de programas de ordenador, sino de un término militar, programar, que significa realizar planes o propuestas de tiempo para el entrenamiento, la logística o el despliegue de las unidades de combate. Dados $n, m \in \mathbb{N}$, la Programación lineal es una rama de la Optimización que se encarga del estudio y resolución de **programas lineales**. Un programa lineal no es mas que un problema general de optimización en el que se consideran restricciones de igualdad y desigualdad de la forma

$$\left.\begin{array}{c} \text{minimizar/maximizar } f = f(\boldsymbol{x}) \text{ sujeto a} \\ f_1(\boldsymbol{x}) = 0, ..., f_k(\boldsymbol{x}) = 0, f_{k+1}(\boldsymbol{x}) \leq 0, ..., f_m(\boldsymbol{x}) \leq 0 \\ \boldsymbol{x} \in \Omega \subset \mathbb{R}^n \end{array}\right\} \quad (\star)$$

donde todas las funciones involucradas son funciones lineales.

Exceptuando al matemático francés G. Monge (1746-1818), quien en 1776 se interesó por problemas de este género, debemos remontarnos al año 1939 para encontrar nuevos

estudios relacionados con los métodos de la actual Programación lineal. Repasamos cronológicamente algunos de los hechos mas relevantes acerca de la aparición de esta teoría.

1. En 1939, el matemático ruso **Leonodas V. Kantarovitch** (1912-1986) publica una extensa monografía titulada *Métodos matemáticos de organización y planificación de la producción* en la que por primera vez se hace corresponder una extensa gama de problemas con una teoría matemática precisa y que dio lugar a la Programación lineal, aunque no se dio a conocer hasta dos décadas mas tarde.

2. Entre 1941 y 1942 se formula por primera vez el *problema de transporte*, estudiado independientemente por **Tjalling C. Koopmans** y **Kantarovitch**. Este tipo de problemas aparecieron al tratar de determinar los planes de embarque al mínimo coste total, conociendo de antemano la disponibilidad y demanda de cada puerto. Ambos fueron galardonados con el Premio Nobel de Economía en 1975, en parte, gracias a su contribución a la Teoría de optimización de recursos.

3. El también premio Nobel en Economía, **George J. Stigler**, publicó en 1946 *Los costes de la subsistencia* donde formula y resuelve el problema de optimización lineal de régimen alimenticio óptimo (Problema de la Dieta).

4. En 1947, **George B. Dantzing**, trabajando como experto en métodos de planificación para las fuerzas aéreas norteamericanas, formula el enunciado general al que se reduce cualquier problema lineal y desarrolla un método iterativo y muy eficaz de resolución, llamado *método del Simplex*. Una de las primeras aplicaciones de sus estudios fue la resolución del llamado "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. Utilizando la Programación lineal, diseñó un plan de abastecimiento aéreo minimizando costes que en pocos meses consiguió igualar a los suministros realizados por carretera y ferrocarril antes del bloqueo. Si bien es cierto que G. Dantzig no fue galardonado con el Premio Nobel, su contribución ha sido reconocida con numerosos premios, entre los que destaca el premio Von Neumann de la Sociedad Americana de Investigación Operativa del año 1975. Durante más de 30 años el método del simplex ha sido el único método utilizado para la resolución de problemas lineales de gran tamaño. Su importancia ha sido tal que en el 2000 fue incluido entre los 10 algoritmos más transcendentes del siglo XX en el "top ten" de la revista *Computing in Science and Engineering*.

5. Por último, en 1947 **John V. Neumann** establece los fundamentos matemáticos de la Programación lineal, al relacionar ésta con la teoría de juegos, que había publicado tres años antes, junto con Oscar Morgenstern, en el libro *Teoría de juegos y comportamiento económico*.

Este trabajo está dedicado a la **Programación no lineal**, que se encarga del estudio y resolución de programas no lineales, a saber, programas del tipo $\star$ pero en el que no se exigen que las funciones involucradas sean lineales. La Programación no lineal tiene tres orígenes muy distintos que nacen en diferentes ramas de la Matemática y por parte de diversos matemáticos como puede verse en el segundo capítulo de este texto, pero todos

culminan en el teorema fundamental de la Programación no lineal, el reconocido Teorema de Karush-Kuhn-Tucker (teorema 9, Cap. 3), que proporciona una condición necesaria para la existencia de mínimos y máximos (locales) en programas no lineales. Como se verá en el último capítulo, bajo hipótesis de convexidad, cualquier extremo local es también global y la condición necesaria provista por el Teorema de Karush-Kuhn-Tucker es también condición suficiente.

En el presente texto, las magnitudes vectoriales serán escritas en letra **negrita** y se utilizará la siguiente notación: $\mathbb{N}$ y $\mathbb{R}$ denotará, respectivamente, el conjunto de los números naturales y de los números reales y $\mathbb{R}_0^+$ y $\mathbb{R}_0^-$ representarán, respectivamente, el conjunto de números reales no negativos y no positivos. Dado $n \in \mathbb{N}$, se denota por $\mathbb{R}^n$ al espacio vectorial euclídeo n-dimensional con el producto escalar usual $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ dado por $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{k=1}^n x_k y_k$ donde $\boldsymbol{x} = (x_1, ..., x_n), \boldsymbol{y} = (y_1, ..., y_n) \in \mathbb{R}^n$. El espacio vectorial de funciones reales y continuas, definidas en un subconjunto abierto $\Omega \subset \mathbb{R}^n$ es denotado por $\mathcal{C}(\Omega)$ y el espacio de funciones reales y derivables k-veces con derivada k-ésima continua, definidas en un subconjunto $\Omega \subset \mathbb{R}^n$ es denotado por $\mathcal{C}^k(\Omega)$. Dados $n, m \in \mathbb{N}$, el espacio vectorial de matrices de tamaño $n \times m$ con coeficientes en $\mathbb{R}$ se representa como $\mathfrak{M}_{n\times m}(\mathbb{R})$ y en el caso particular en el que $n = m$ (matrices cuadradas), se escribirá simplemente $\mathfrak{M}_n(\mathbb{R})$. La matriz traspuesta de una matriz $M \in \mathfrak{M}_{n\times m}(\mathbb{R})$ será denotada por M^T. Si $\Omega \subset \mathbb{R}^n$ es un subconjunto de $\mathbb{R}^n$, entonces se escribirá $\overline{\Omega}$ para la clausura de Ω y Ω° para el interior de Ω. Los símbolos $\sum$ y $\prod$ denotarán, respectivamente, la suma y el producto de los términos indicados.
Si $f = f(x_1, ..., x_n) : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ es una función suficientemente diferenciable se denota, por ejemplo

$$D_k f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial x_k}, \quad D_{ijk} f(\boldsymbol{x}) = \frac{\partial^3 f(\boldsymbol{x})}{\partial x_i \partial x_j \partial x_k} \quad \text{o} \quad D_{iijj} f(\boldsymbol{x}) = \frac{\partial^4 f(\boldsymbol{x})}{\partial x_i^2 \partial x_j^2}$$

para ciertos índices $i, j, k \in \{1, ..., n\}$. El símbolo $\nabla f(\boldsymbol{x})$ denotará el vector gradiente de una función $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ diferenciable en un punto $\boldsymbol{x} \in \Omega$, es decir,

$$\nabla f(\boldsymbol{x}) = (D_1 f(\boldsymbol{x}), ..., D_n f(\boldsymbol{x}))$$

La norma euclídea en $\mathbb{R}^n$ será representada como $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$, esto es,

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

donde $\boldsymbol{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ y la bola abierta (en $\mathbb{R}^n$) de centro $\boldsymbol{a} \in \mathbb{R}^n$ y radio $\rho \in \mathbb{R}^+$, será escrita como $B(\boldsymbol{a}, \rho) = \{\boldsymbol{x} \in \mathbb{R}^n : \ \|\boldsymbol{x} - \boldsymbol{a}\| < \rho\}$.
Dados $n, m \in \mathbb{N}$ y una función diferenciable $\boldsymbol{f} = (f_1, ..., f_m) : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, la matriz jacobiana de $\boldsymbol{f}$ en el punto $\boldsymbol{x}^* \in \Omega$ será escrito como

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

Dados $n \in \mathbb{N}$ y una función $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ de clase $\mathcal{C}^2$ en Ω, la matriz hessiana de f en el punto $\boldsymbol{x}^* \in \Omega$ será escrita como

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

Finalmente, me gustaría mencionar algunos resultados del Análisis matemático que estarán presentes en muchas ocasiones a lo largo del trabajo, por lo que es conveniente que el lector este familiarizado con ellos.

Teorema de Weiertrass: ver capítulo 4 de [5].

Fórmula del resto de Taylor: ver capítulo 5 de [5].

Formas cuadráticas y clasificación: ver capítulo 1 de [40].

Teorema de Bolzano-Weiertrass: véase [21].

Teorema de conservación del signo: véase [38].

Resumen

El presente Trabajo de fin de grado titulado *Una generalización del Teorema de los multiplicadores de Lagrange: condiciones de Karush-Kuhn-Tucker en Programación no lineal* consta de los siguientes tres capítulos:

- Conocimientos previos sobre extremos libres y los multiplicadores de Lagrange
- Orígenes históricos de la Programación no lineal.
- El Teorema de Karush-Kuhn-Tucker y aplicaciones en Programación convexa.

Los objetivos a alcanzar son los siguientes:

(1) Resumen de lo estudiado a lo largo del grado sobre optimización incluyendo numerosos ejemplos.

(2) Origen y contexto histórico de la Programación no lineal y, en especial, de su teorema fundamental: el Teorema de Karush-Kuhn-Tucker.

(3) Demostración del Teorema de Karush-Kuhn-Tucker y ejemplos.

(4) Resultados principales acerca de conjuntos y funciones convexas.

(5) Programación convexa, condición suficiente de existencia de solución en programas convexos y ejemplos.

A continuación, se resume el contenido de cada uno de los capítulos.

Capítulo 1: *Resumen del método de los multiplicadores de Lagrange.*
Este capítulo cubre el objetivo (1) y se divide en dos partes. En la primera parte se estudian condiciones necesarias y condiciones suficientes para la existencia de extremos libres, distinguiendo entre extremos locales y globales, de funciones reales de una y varias variables. Ambos casos, son ilustrados con ejemplos concretos. Esta primera parte corresponde a las secciones 1.1 y 1.2.
La segunda parte está dedicada al clásico método de los multiplicadores de Lagrange, que sirve como una pequeña aproximación al Teorema de Karush-Kuhn-Tucker, que lo incluye como caso particular y que trata sobre la búsqueda de extremos condicionados sujetos a restricciones de igualdad. Aquí se hablará un poco de historia sobre el matemático J. L. Lagrange y el origen de su método, se describirán los pasos a seguir para la correcta aplicación de tal método y, de nuevo, será ilustrado con numerosos ejemplos. Esta segunda parte corresponde a la sección 1.3.

Capítulo 2. *Orígenes históricos de la Programación no lineal.*
Este capítulo cubre el objetivo (2). En este capítulo, tras una pequeña introducción y unos prerrequisitos matemáticos (sección 2.1), se cuentan básicamente los tres orígenes tan dispersos que tuvo el Teorema de Karush-Kuhn-Tucker: origen en el Cálculo de variaciones por parte de W. Karush (sección 2.2), origen en la teoría de la convexidad por parte de F. John (sección 2.3) y origen en la Programación lineal por parte del alumno y profesor en la Universidad de Princeton, Kuhn y Tucker, respectivamente (sección 2.4). Finalmente, se habla sobre la importancia del contexto histórico y social y de cómo puede influir estos a la hora de clasificar un descubrimiento matemático como múltiple o no (secciones 2.5 y 2.6). El capítulo entero es un amplio resumen del artículo [2].

Capítulo 3. *El Teorema de Karush-Kuhn-Tucker y aplicaciones en Programación convexa.*
Este capítulo cubre los objetivos restantes (3)-(6). En este capítulo se diferencian cuatro partes:
En la primera parte se dan las nociones fundamentales para trabajar en Programación mixta, entre las que destacamos el concepto de regularidad de un punto (defición 16), seguidamente del Teorema de Karush-Kuhn-Tucker que nos provee de una condición necesaria para la existencia de mínimo local en un programa con restricciones mixtas (teorema 9); así como, de la correspondiente versión para máximo local (corolario 1) que se deduce directamente del mencionado teorema. La demostración que aquí se hace de este teorema no es la más habitual en la comunidad matemática pero sí, quizás, una de la más simples por el hecho de usar resultados básicos del Análisis matemático y se debe al matemático americano E. J. McShane. Al igual que se hizo para el método de los multiplicadores de Lagrange, también se comentan los pasos a seguir para resolver un programa mixto en la práctica y se ilustra este procedimiento con ejemplos (sección 3.1). Esta primera parte barre el objetivo (3).
En la segunda parte se introducen los conjuntos convexos y las funciones convexas y cóncavas, dándose las principales propiedades y caracterizaciones sobre estos y haciendo ver que las funciones convexas y cóncavas juegan un papel relevante en Optimización (teorema 14 y corolario 5). La finalidad de esta parte es familiarizarnos con este tipo de funciones

con vistas a la última parte del capítulo, en las que estarán presentes constantemente. Además, se incluyen varios ejemplos y comentarios tanto para conjuntos convexos, como para funciones convexas y cóncavas, a lo largo de esta parte del trabajo (sección 3.2). Esta segunda parte barre el objetivo (4).
En la tercera parte se introduce lo que se entiendo por un programa convexo y cóncavo, donde surge un nuevo concepto: la superconsistencia (definición 20) y se demuestra que, bajo hipótesis de convexidad (resp. concavidad) y superconsistencia o regularidad, la condición necesaria para la existencia de mínimo (resp. máximo) local de un programa convexo (resp. cóncavo) que aporta el Teorema de Karush-Kuhn-Tucker es también condición suficiente (teorema 15, resp. corolario 7), (sección 3.3). Esta tercera parte barre el objetivo (5).
Finalmente, aunque no sea propiamente un objetivo a cumplir, en el apartado 3.4 se hace una breve presentación sobre Programación convexa dual: se motiva la definición de programa dual, se prueba su principal resultado (teorema 17) que relaciona la solución de un programa convexo con la de su programa dual, lo que permite resolver un programa convexo resolviendo su correspondiente dual asociado, que por lo general suele ser menos complicado. También se describen los pasos a seguir para realizar todo este proceso correctamente y se aclara con varios ejemplos (sección 3.4). Por último, se incluyen tres casos particulares de Programación convexa, a saber, Programación lineal, geométrica y cuadrática, que se pueden atacar satisfactoriamente a través de la dualidad. La principal fuente consultada en este capítulo ha sido [40].

Summary

The present end degree work entitled *A generalization of the Lagrange multiplier rule: Karush-Kuhn-Tucker conditions in nonlinear Programming* consists of the following chapters:

- Previous knowledge of the free extreme points of a real function and the Lagrange multiplier rule.
- Historical origins of nonlinear Programming.
- The Karush-Kuhn-Tucker Theorem and applications to convex Programming.

The objectives to achieve here are the followings:

(1) To summarize the main results and ideas shown, along the Degree, and related to Optimization.

(2) Historical origin and context of the nonlinear Programming and, in special, of his main theorem: the Karush-Kuhn-Tucker Theorem.

(3) Proof of the Karush-Kuhn-Tucker Theorem and examples.

(4) Main results about convex sets and functions.

(5) Convex Programming, sufficient condition for the existence of solution in a convex program and examples.

Now, the content of these three chapters is summarized below.

Chapter 1. *Summary of the Lagrange multiplier rule.*
This chapter reach the target (1) and it's divided into two differents parts. The first one, is devoted to the neccesary conditions and sufficient conditions for the existence of free minimizers and maximizers, distinguishing between globals or locals minimizers and maximizers, of one real variable and several real variables functions. Both cases are illustrated with examples. This first parte correspond to the sections 1.1 and 1.2.
The second one is devoted to the classical Lagrange multiplier rule, which acts as a first approximation to the Karush-Kuhn-Tucker Theorem, (a generalization of it) and deals with the search for conditional extremes when only equality constraints are presented. Here, we will talk about a little bit of history of Lagrange and the origin of his method. We will enumerate the steps to be taken to apply correctly his rule and, again, it will be illustrated with various examples. This second part correspond to the section 1.3.

Chapter 2. *Historical origins of nonlinear Programming.*
This chapter reach the target (2). In this chapter, after a brief introduction and some mathematical prerequisites (section 2.1), we include basically the so dispersed three origins of the Karush-Kuhn-Tucker Theorem: the origin in the Calculus of variations on the part of W. Karush (section 2.2), the origin in the theory of convexity on the part of F. John (section 2.3) and the proper origin in the linear Programming on the part of the student and teacher from the University of Princeton, Kuhn y Tucker, respectively (section 2.4). Finally, we talk about the importance of the hitorical and social context and about how it can affect when it comes to classify a mathematical discover as a multiples or singletons, instead (sections 2.5 and 2.6). The hole chapter is an extensive summary of the article [2] by T. H. Kjeldsen.

Chapter 3. *The Karush-Kuhn-Tucker Theorem and applications to convex Programming.*
This chapter reach the rest targets (3)-(5). This chapter is divided into four parts:
In the first one, we will give the fundamental definitions to work in programs subject to both equality and inequality constraints, which are called mixed programs, among them regular point (definition 16) and then the proof of the main theorem of this work: the Karush-Kuhn-Tucker Theorem, which provides neccesary conditions for a local minimizer for a mixed program (theorem 9) and the appropiate version for the local maximizer (corolario 1), which is a direct consecuence of theorem 9. The proof we did here is not the most common one in the mathematical literature but it is one of the most simple because it uses basically elemetary results from Mathematical Analysis and it's due to the American mathematician E. J. McShane. Just like it was done with the Lagrange multiplier rule, we enumerate the steps to solve correctly a mixed program, together with some examples (section 3.1). This first part reach the objetive (3).
In the second part, convex sets and convex and concave functions are presented, giving the main properties and characterizations of them and bringing to light the fact that the convex and concave functions play an important role in Optimization (theorem 14 and corollary 5). The goal of this part is to become comfortable with this kind of functions with a view to the last part of the chapter, where they will be presented constantly. Futhermore, it also includes several examples and notes both for convex sets and for convex

and concave functions, along the section 3.2. This second parte reach the objetive (4).
In the third part, we define a convex and concave program, where a new concept arises: the superconsistence (definition 20) and we proove that, assumming some hypothesis related to convexity (resp. concavity) and superconsistence or regularity, the necessary condition for the existence of a local minimizer (resp. maximizer) of a convex program given by the Karush-Kuhn-Tucker Theorem, is also a sufficient condition (theorem 15, resp. 7), (section 3.3). This third part reach the objetive (5).
Finally, although it is not a proper objetive of this work, we describe a brief introduction to the dual convex Programming: we motivate the definition of a dual program, we proove its main result (theorem 17) which relates the solutions of a convex program and its associated dual program. This says that we can solve a convex program by solving its dual program (which often it is easier). To do that, we describe this procedure step by step and we illustrate it with some examples (section 3.4). At the end, we take into account some particular cases of convex Programming, namely linear, geometric and quadratic Program, which can be attacked successfully using the duality. The main consulted source in this chapter was [40].

Capítulo 1

Conocimientos previos sobre extremos libres y los multiplicadores de Lagrange

En este capítulo se exponen resumidamente los principales resultados, relacionados con los estudiados en el grado de Matemáticas, sobre el cálculo de extremos libres y condicionados, distinguiendo en ambos casos tanto extremos locales como globales. En lo que sigue, se enuncian las condiciones necesarias y las condiciones suficientes para la existencia de extremos (locales y globales) de una función de una y varias variables reales, acompañados con ejemplos. Finalmente, se enuncia el teorema que establece la validez del clásico *método de los multiplicadores de Lagrange* para el cálculo de extremos de una función de varias variables reales condicionados a ciertas restricciones de igualdad y se describe brevemente este método junto con numerosos ejemplos que ilustrarán su gran potencial.

1.1. Extremos libres para funciones de una variable

Comenzamos esta sección con las definiciones elementales de *extremo global y local*, *punto crítico* y *punto de silla* de una función real de variable real.

Definición 1 (Extremos globales)**.** *Sea $A \subseteq \mathbb{R}$ un subconjunto no vacío de $\mathbb{R}$, $a \in A$ un punto de A y $f : A \to \mathbb{R}$ una función real definida en A, entonces*

- *f tiene un **máximo global** en a si $f(a) \geq f(x)$ para todo $x \in A$.*
- *f tiene un **máximo global estricto** en a si $f(a) > f(x)$ para todo $x \in A \setminus \{a\}$, es decir, que a es el único punto de A donde f tiene un máximo global.*
- *f tiene un **mínimo global** en a si $f(a) \leq f(x)$ para todo $x \in A$.*
- *f tiene un **mínimo global estricto** en a si $f(a) < f(x)$ para todo $x \in A \setminus \{a\}$, es decir, que a es el único punto de A donde f tiene un mínimo global.*

*Las expresiones **extremo global** o extremo absoluto se usan para referirse indistintamente a un máximo o mínimo global.*

Definición 2 (Extremos locales)**.** *Sea $A \subseteq \mathbb{R}$ un subconjunto no vacío de $\mathbb{R}$, $a \in A$ un punto de A y $f : A \to \mathbb{R}$ una función real definida en A, entonces*

- *f tiene un **máximo local** en a si existe un entorno abierto $N(a) \subset A$ de a en A tal que $f(a) \geq f(x)$ para todo $x \in N(a)$.*
- *f tiene un **máximo local estricto** en a si existe un entorno abierto $N(a) \subset A$ de a en A tal que $f(a) > f(x)$ para todo $x \in N(a) \setminus \{a\}$, es decir, que a es el único punto de $N(a)$ donde f tiene un máximo local.*
- *f tiene un **mínimo local** en a si existe un entorno abierto $N(a) \subset A$ de a en A tal que $f(a) \leq f(x)$ para todo $x \in N(a)$.*
- *f tiene un **mínimo local estricto** en a si existe un entorno abierto $N(a) \subset A$ de a en A tal que $f(a) < f(x)$ para todo $x \in N(a) \setminus \{a\}$, es decir, que a es el único punto de $N(a)$ donde f tiene un mínimo local.*

*Las expresiones **extremo local** o extremo relativo se usan para referirse indistintamente a un máximo o mínimo local.*

Además, en la situación de la definición anterior, es claro que una función f tiene un máximo (resp. mínimo) local o global (estricto) en un punto a si, y sólo si, $-f$ tiene un mínimo (resp. máximo) local o global (estricto) en a.

Definición 3 (Punto crítico)**.** *Sea $A \subseteq \mathbb{R}$ un subconjunto no vacío de $\mathbb{R}$ y $f : A \to \mathbb{R}$ una función real definida en A, entonces un **punto crítico** de f es un punto $a \in A^\circ$ donde f es derivable con $f'(a) = 0$.*

Definición 4 (Punto de silla)**.** *Sea $A \subseteq \mathbb{R}$ un subconjunto no vacío de $\mathbb{R}$ y $f : A \to \mathbb{R}$ una función real definida en A, entonces un **punto de silla** de f es un punto crítico de f donde f no tiene ningún valor extremo.*

A continuación distinguimos dos partes en esta sección: una trata sobre el cálculo de extremos locales y la otra trata sobre el cálculo de extremos globales de funciones reales de una variable.

1.1.1. Extremos locales libres

En lo que sigue se enuncian condiciones necesarias y condiciones suficientes de existencia de extremo local para funciones reales de variable real.

Teorema 1 (Condición necesaria)**.** *Sea $I \subseteq \mathbb{R}$ un intervalo no trivial de $\mathbb{R}$, $x_0 \in I^\circ$ un punto interior de I y $f : I \to \mathbb{R}$ una función real definida en I. Si f es derivable en x_0 y tiene un extremo local en x_0, entonces x_0 es un punto crítico de f, esto es, la derivada de f en x_0 es cero.*

La demostración de este teorema puede encontrarse en el capítulo 5 de [5] y utiliza el hecho de que si $f'(x_0) > 0$ o $f'(x_0) = +\infty$ (resp. $f'(x_0) < 0$ o $f'(x_0) = -\infty$), entonces existe un entorno abierto $N(x_0)$ de x_0 en I de manera que las diferencias $f(x) - f(x_0)$ y $x - x_0$ tienen el mismo signo (resp. el signo opuesto) para todo $x \in N(x_0)$. Así, si existe

$f'(x_0)$ ha de ser 0 pues, en caso contrario, f no tendría ni máximo ni mínimo en x_0.

El recíproco del teorema 1 no es cierto, como puede comprobarse con $f(x) = x^3$ definida en todo $\mathbb{R}$ y $x_0 = 0$. En este caso, $f'(0) = 0$ pero f toma valores positivos y negativos en cualquier entorno del origen.

Intuitivamente, el estudio del cambio de la monotonía de una función en un cierto punto proporciona una condición suficiente para que una función derivable alcance un extremo local en dicho punto.

Teorema 2 (Condición suficiente)**.** *Sea I un intervalo no trivial de $\mathbb{R}$, $x_0 \in I^\circ$ un punto interior de I y $f : I \to \mathbb{R}$ una función real definida en I. Si f es continua en todo I y derivable en un entorno de x_0, salvo quizás en x_0, entonces*

- *si existe $\delta > 0$ tal que $f'(x) > 0$ para todo $x \in (x_0 - \delta, x_0)$ y $f'(x) < 0$ para todo $x \in (x_0, x_0 + \delta)$, entonces f tiene un máximo local en x_0.*

- *si existe $\delta > 0$ tal que $f'(x) < 0$ para todo $x \in (x_0 - \delta, x_0)$ y $f'(x) > 0$ para todo $x \in (x_0, x_0 + \delta)$, entonces f tiene un mínimo local en x_0.*

- *si existe $\delta > 0$ tal que f' mantiene el signo constante en $(x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$, entonces f no alcanza ningún extremo local en x_0.*

Aunque, en la práctica, el resultado más usado para garantizar que una función real de una variable alcanza un extremo local en un punto es el siguiente.

Teorema 3 (Condición suficiente)**.** *Sea $m \in \mathbb{N}$ un número natural con $m \geq 2$, $I \subseteq \mathbb{R}$ un intervalo no trivial de $\mathbb{R}$, $x_0 \in I^\circ$ un punto interior de I y $f : I \to \mathbb{R}$ una función real de clase $\mathcal{C}^m$ definida en I. Si $f'(x_0) = \cdots = f^{(m-1)}(x_0) = 0$ y $f^{(m)}(x_0) \neq 0$, entonces*

- *si m es par y $f^{(m)}(x_0) > 0$, f tiene un mínimo local (estricto) en x_0.*

- *si m es par y $f^{(m)}(x_0) < 0$, f tiene un máximo local (estricto) en x_0.*

- *si m es impar, f tiene un punto de silla de x_0.*

La demostración de este teorema puede encontrarse en el capítulo 7 de [5] y se deduce de la Fórmula de Taylor para funciones de una variable.

Dados $a, b \in \mathbb{R}$ con $a < b$, $I = [a, b]$ un intervalo cerrado y acotado de $\mathbb{R}$ y $f : I \to \mathbb{R}$ una función real y continua definida en I, el Teorema de Weiertrass garantiza que f tiene mínimo y máximo global en I y por el teorema 1, los únicos puntos donde f puede tener un valor extremo son los puntos críticos, los puntos de I donde f no es derivable y los puntos de la frontera de I, a saber, a y b.

1.1.2. Extremos globales libres

En lo que sigue se enuncian condiciones necesarias y condiciones suficientes de existencia de extremo global para funciones reales de variable real y se ilustran con tres ejemplos concretos.

La condición necesaria de extremo global es idéntica a la de extremo local (teorema 1) y la condición suficiente de extremo global es la siguiente.

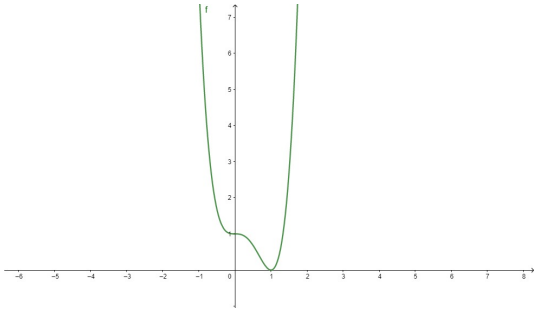
Teorema 4 (Condición suficiente)**.** *Sea $m \in \mathbb{N}$ un número natural con $m \geq 2$, $I \subseteq \mathbb{R}$ un intervalo no trivial de $\mathbb{R}$, $x_0 \in I^\circ$ un punto interior de I y $f : I \to \mathbb{R}$ una función real de clase $\mathcal{C}^m$ definida en I. Si $f'(x_0) = \cdots = f^{(m-1)}(x_0) = 0$ y $f^{(m)}(x_0) \neq 0$ y m es par, entonces*

- *si $f^{(m)}(x) \geq 0$ para todo $x \in I$, f tiene un mínimo global en x_0.*
- *si $f^{(m)}(x) > 0$ para todo $x \in I \setminus \{x_0\}$, f tiene un mínimo global estricto en x_0.*
- *si $f^{(m)}(x) \leq 0$ para todo $x \in I$, f tiene un máximo global en x_0.*
- *si $f^{(m)}(x) < 0$ para todo $x \in I \setminus \{x_0\}$, f tiene un máximo global estricto en x_0.*

La demostración de este teorema puede encontrarse en el capítulo 1 de [40] y se deduce de la Fórmula de Taylor para funciones de una variable.

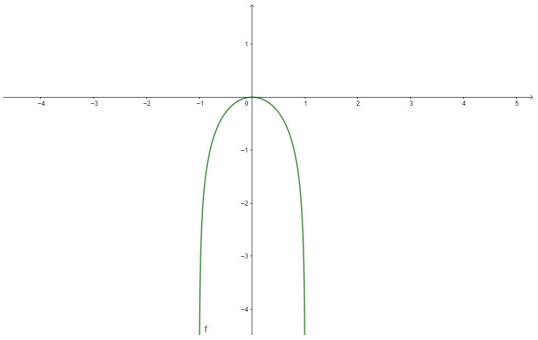
A continuación, se expone un ejemplo donde puede apreciarse la aplicabilidad de los resultados anteriores.

Ejemplo 1. *Considérese la función $f(x) = 3x^4 - 4x^3 + 1$ definida en todo $\mathbb{R}$. Como $f'(x) = 12x^2(x-1)$ para todo $x \in \mathbb{R}$, los únicos puntos críticos de f son $x = 0$ y $x = 1$. Como $f''(x) = 12x(3x-2)$ para todo $x \in \mathbb{R}$ y $f''(0) = 0$ y $f''(1) = 12$, en virtud del teorema 3, se sigue que f tiene un mínimo local en $x = 1$ y como $f'''(x) = 72x - 24$ para todo $x \in \mathbb{R}$ y $f'''(0) = -24$, de nuevo por el teorema 3, f tiene un punto de silla en $x = 0$.*



Nótese que $\lim_{x\to\pm\infty} f(x) = +\infty$, luego f no tiene máximo global en $\mathbb{R}$. Además, el mínimo local $x = 1$ es también mínimo global de f en $\mathbb{R}$.

Ejemplo 2. *Considérese la función $f(x) = \log(1-x^2)$ definida en el intervalo $I = (-1, 1)$. Como $f'(x) = -\frac{2x}{1-x^2}$ para todo $x \in I$, el único punto crítico de f es $x = 0$. Como $f''(x) = -\frac{2(1+x^2)}{(1-x^2)^2} < 0$ para todo $x \in I$, en virtud del teorema 4, se sigue que f tiene un máximo global estricto en $x = 0$.*



Nótese que $\lim_{x\to\pm 1} f(x) = -\infty$*, luego* f *no tiene mínimo global en* I*.*

Ejemplo 3. *La función* $f : \mathbb{R} \to \mathbb{R}$ *dada por* $f(x) = e^{-x}\sin(x)$ *para todo* $x \in \mathbb{R}$ *es de clase dos en* $\mathbb{R}$ *con* $f'(x) = e^{-x}(\cos(x) - \sin(x))$ *y* $f''(x) = -2e^{-x}\cos(x)$ *para todo* $x \in \mathbb{R}$*. Los puntos críticos de* f *son* $\pi/4 + 2\pi k$ *y* $5\pi/4 + 2\pi k$ *para cada* $k \in \mathbb{Z}$ *y como* $f''(\pi/4 + 2\pi k) < 0$ *y* $f''(5\pi/4 + 2\pi k) > 0$ *para cada* $k \in \mathbb{Z}$*, en virtud del teorema 3, se sigue que* f *tiene un máximo local en cada punto de la forma* $x_k = \pi/4 + 2\pi k$ *y un mínimo local en cada punto de la forma* $x_k = 5\pi/4 + 2\pi k$ *con* $k \in \mathbb{Z}$*.*

1.2. Extremos libres para funciones de varias variables

Comenzamos esta sección con las definiciones elementales de *extremo global y local*, *punto crítico* y *punto de silla* de una función real de varias variables reales.

Definición 5 (Extremos globales)**.** *Sea* $n \in \mathbb{N}$ *un número natural,* $A \subseteq \mathbb{R}^n$ *un subconjunto no vacío de* $\mathbb{R}^n$*,* $\boldsymbol{a} \in A$ *un punto de* A *y* $f : A \to \mathbb{R}$ *una función real definida en* A*, entonces*

- f *tiene un* ***máximo global*** *en* $\boldsymbol{a}$ *si* $f(\boldsymbol{a}) \geq f(\boldsymbol{x})$ *para todo* $\boldsymbol{x} \in A$*.*
- f *tiene un* ***máximo global estricto*** *en* $\boldsymbol{a}$ *si* $f(\boldsymbol{a}) > f(\boldsymbol{x})$ *para todo* $\boldsymbol{x} \in A \setminus \{\boldsymbol{a}\}$*, es decir, que* $\boldsymbol{a}$ *es el único punto de* A *donde* f *tiene un máximo global.*
- f *tiene un* ***mínimo global*** *en* $\boldsymbol{a}$ *si* $f(\boldsymbol{a}) \leq f(\boldsymbol{x})$ *para todo* $\boldsymbol{x} \in A$*.*
- f *tiene un* ***mínimo global estricto*** *en* $\boldsymbol{a}$ *si* $f(\boldsymbol{a}) < f(\boldsymbol{x})$ *para todo* $\boldsymbol{x} \in A \setminus \{\boldsymbol{a}\}$*, es decir, que* $\boldsymbol{a}$ *es el único punto de* A *donde* f *tiene un mínimo global.*

Las expresiones ***extremo global*** *o extremo absoluto se usan para referirse indistintamente a un máximo o mínimo global.*

Definición 6 (Extremos locales)**.** *Sea* $n \in \mathbb{N}$ *un número natural,* $A \subseteq \mathbb{R}^n$ *un subconjunto no vacío de* $\mathbb{R}^n$*,* $\boldsymbol{a} \in A$ *un punto de* A *y* $f : A \to \mathbb{R}$ *una función real definida en* A*, entonces*

- f *tiene un* ***máximo local*** *en* $\boldsymbol{a}$ *si existe un entorno abierto* $N(\boldsymbol{a}) \subset A$ *de* $\boldsymbol{a}$ *en* A *tal que* $f(\boldsymbol{a}) \geq f(\boldsymbol{x})$ *para todo* $\boldsymbol{x} \in N(\boldsymbol{a})$*.*

- *f tiene un **máximo local estricto** en $\boldsymbol{a}$ si existe un entorno abierto $N(\boldsymbol{a}) \subset A$ de $\boldsymbol{a}$ en A tal que $f(\boldsymbol{a}) > f(\boldsymbol{x})$ para todo $\boldsymbol{x} \in N(\boldsymbol{a}) \setminus \{\boldsymbol{a}\}$, es decir, que $\boldsymbol{a}$ es el único punto de $N(\boldsymbol{a})$ donde f tiene un máximo local.*
- *f tiene un **mínimo local** en $\boldsymbol{a}$ si existe un entorno abierto $N(\boldsymbol{a}) \subset A$ de $\boldsymbol{a}$ en A tal que $f(\boldsymbol{a}) \leq f(\boldsymbol{x})$ para todo $\boldsymbol{x} \in N(\boldsymbol{a})$.*
- *f tiene un **mínimo local estricto** en $\boldsymbol{a}$ si existe un entorno abierto $N(\boldsymbol{a}) \subset A$ de $\boldsymbol{a}$ en A tal que $f(\boldsymbol{a}) < f(\boldsymbol{x})$ para todo $\boldsymbol{x} \in N(\boldsymbol{a}) \setminus \{\boldsymbol{a}\}$, es decir, que $\boldsymbol{a}$ es el único punto de $N(\boldsymbol{a})$ donde f tiene un mínimo local.*

*Las expresiones **extremo local** o extremo relativo se usan para referirse indistintamente a un máximo o mínimo local.*

Además, en la situación de la definición anterior, es claro que una función f tiene un máximo (resp. mínimo) local o global (estricto) en un punto $\boldsymbol{a}$ si, y sólo si, $-f$ tiene un mínimo (resp. máximo) local o global (estricto) en $\boldsymbol{a}$.

Definición 7 (Punto crítico)**.** *Sea $n \in \mathbb{N}$ un número natural, $A \subseteq \mathbb{R}^n$ un subconjunto no vacío de $\mathbb{R}^n$ y $f : A \to \mathbb{R}$ una función real definida en A, entonces un **punto crítico** de f es un punto $\boldsymbol{a} \in A^\circ$ donde f es diferenciable con $\nabla f(\boldsymbol{a}) = 0$.*

Definición 8 (Punto de silla)**.** *Sea $n \in \mathbb{N}$ un número natural, $A \subseteq \mathbb{R}^n$ un subconjunto no vacío de $\mathbb{R}^n$ y $f : A \to \mathbb{R}$ una función real definida en A, entonces un **punto de silla** de f es un punto crítico de f donde f no tiene ningún valor extremo.*

A continuación distinguimos dos partes en esta sección: una trata sobre el cálculo de extremos locales y la otra trata sobre el cálculo de extremos globales de funciones reales de varias variables reales.

1.2.1. Extremos locales libres

En lo que sigue se enuncian condiciones necesarias y condiciones suficientes de existencia de extremo local para funciones reales de varias variables reales.

Teorema 5 (Condición necesaria)**.** *Sea $n \in \mathbb{N}$ un número natural, $A \subseteq \mathbb{R}^n$ un subconjunto no vacío de $\mathbb{R}^n$, $\boldsymbol{x_0} \in A^\circ$ un punto interior de A y $f : A \to \mathbb{R}$ una función real definida en A. Si f es diferenciable en $\boldsymbol{x_0}$ y tiene un extremo local en $\boldsymbol{x_0}$, entonces $\boldsymbol{x_0}$ es un punto crítico de f, esto es, el gradiente de f en $\boldsymbol{x_0}$ es el vector cero.*

La demostración de este teorema puede encontrarse en el capítulo 1 de [40] y utiliza la correspondiente condición necesaria para funciones de una variable (teorema 1) aplicadas a las funciones derivables (de una variable) $g_k(x) = f(x_1^0, ..., x_{k-1}^0, x, x_{k+1}^0, ..., x_n^0)$ con $k = 1, ..., n$, definidas en un entorno abierto de x_k^0 y siendo $(x_1^0, ..., x_n^0)$ el extremo local de f.

El recíproco del teorema 5 no es cierto, como puede comprobarse con $f(x, y) = (y - x^2)(y - 2x^2)$ definida en todo $\mathbb{R}^2$ y $\boldsymbol{x_0} = (0, 0)$. En este caso, $D_1 f(0, 0) = 0$ y $D_2 f(0, 0) = 0$ pero f toma valores positivos y negativos en cualquier entorno del origen.

La condición suficiente de extremo local requiere el conocimiento previo del concepto de forma cuadrática y su clasificación y es la siguiente.

Teorema 6 (Condición suficiente). *Sea $n \in \mathbb{N}$ un número natural, $A \subseteq \mathbb{R}^n$ un subconjunto no vacío de $\mathbb{R}^n$, $\boldsymbol{x_0} \in A^\circ$ un punto interior de A y $f : A \to \mathbb{R}$ una función real de clase $\mathcal{C}^2$ definida en A. Si $\boldsymbol{x_0}$ es un punto crítico de f, entonces*

- *si $H_f(\boldsymbol{x_0})$ es definida positiva, f tiene un mínimo local (estricto) en $\boldsymbol{x_0}$.*
- *si $H_f(\boldsymbol{x_0})$ es definida negativa, f tiene un máximo global (estricto) en $\boldsymbol{x_0}$.*
- *si $H_f(\boldsymbol{x_0})$ es indefinida, f tiene un punto de silla en $\boldsymbol{x_0}$.*

La demostración de este teorema puede encontrarse en el capítulo 7 de [5] y se deduce de la Fórmula de Taylor para funciones de varias variables.

Dados $n \in \mathbb{N}$ un número natural, $A \subset \mathbb{R}^n$ un subconjunto compacto de $\mathbb{R}^n$ y $f : A \to \mathbb{R}$ una función real y continua definida en A, el Teorema de Weiertrass garantiza que f tiene mínimo y máximo global en A y por el teorema 5, los únicos puntos donde f puede tener un valor extremo son los puntos críticos, los puntos de A donde f no es diferenciable y los puntos de la frontera de A.

1.2.2. Extremos globales libres

En lo que sigue se enuncian condiciones necesarias y condiciones suficientes de existencia de extremo global para funciones reales de varias variables reales y se ilustran con tres ejemplos concretos.

La condición necesaria de extremo global es idéntica a la de extremo local (teorema 5). De nuevo, la condición suficiente de extremo global requiere el conocimiento previo del concepto de forma cuadrática y su clasificación y es la siguiente.

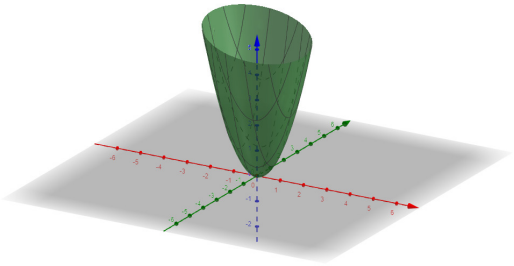
Teorema 7 (Condición suficiente). *Sea $n \in \mathbb{N}$ un número natural, $A \subseteq \mathbb{R}^n$ un subconjunto no vacío de $\mathbb{R}^n$, $\boldsymbol{x_0} \in A^\circ$ un punto interior de A y $f : A \to \mathbb{R}$ una función real de clase $\mathcal{C}^2$ definida en A. Si $\boldsymbol{x_0}$ es un punto crítico de f, entonces*

- *si H_f es semidefinida positiva en A, f tiene un mínimo global en $\boldsymbol{x_0}$.*
- *si H_f es definida positiva en A, f tiene un mínimo global estricto en $\boldsymbol{x_0}$.*
- *si H_f es semidefinida negativa en A, f tiene un máximo global en $\boldsymbol{x_0}$.*
- *si H_f es definida negativa en A, f tiene un máximo global estricto en $\boldsymbol{x_0}$.*

La demostración de este teorema puede encontrarse en el capítulo 1 de [40] y se deduce de la Fórmula de Taylor para funciones de varias variables.

A continuación, se expone un ejemplo donde puede apreciarse la aplicabilidad de los resultados anteriores.

Ejemplo 4. *Considérese la función* $f(x,y) = x^2 + y^2 + xy$ *definida en todo* $\mathbb{R}^2$. *Como* $\frac{\partial f}{\partial x}(x,y) = 2x + y$ *y* $\frac{\partial f}{\partial y}(x,y) = 2y + x$ *para todo* $(x,y) \in \mathbb{R}^2$, *el único punto crítico de* f *es* $(x,y) = (0,0)$ *y como* $H_f(x,y) = H_f(0,0) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ *para todo* $(x,y) \in \mathbb{R}^2$ *es una matriz definida positiva, en virtud del teorema 7, se sigue que* f *tiene un mínimo global estricto en* $(0,0)$.

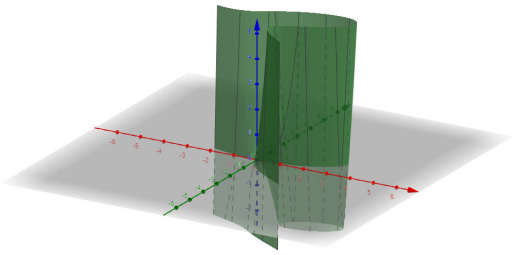


Nótese que $\lim_{x,y\to\infty} f(x,y) = +\infty$, *luego* f *no tiene máximo global en* $\mathbb{R}$.

Ejemplo 5. *Considérese la función* $f(x,y) = x^3 - 12xy + 8y^3$ *definida en todo* $\mathbb{R}^2$. *Como* $\frac{\partial f}{\partial x}(x,y) = 3x^2 - 12y$ *y* $\frac{\partial f}{\partial y}(x,y) = -12x + 24y^2$ *para todo* $(x,y) \in \mathbb{R}^2$, *los únicos puntos críticos de* f *son* $(x,y) = (2,1)$ *y* $(x,y) = (0,0)$ *y como* $H_f(x,y) = \begin{pmatrix} 6x & -12 \\ -12 & 48y \end{pmatrix}$ *para todo* $(x,y) \in \mathbb{R}^2$, *entonces*

$$H_f(2,1) = \begin{pmatrix} 12 & -12 \\ -12 & 48 \end{pmatrix} \qquad y \qquad H_f(0,0) = \begin{pmatrix} 0 & -12 \\ -12 & 0 \end{pmatrix}$$

Como $H_f(2,1)$ *es una matriz definida positiva, se sigue del teorema 6 que* f *tiene un mínimo local estricto en* $(2,1)$.
Como $H_f(0,0)$ *es una matriz indefinida, se sigue del teorema 6 que* f *tiene un punto de silla en* $(0,0)$.



Nótese que $\lim_{x\to\pm\infty} f(x,0) = \pm\infty$, *luego* f *no tiene ni máximo ni mínimo global en* $\mathbb{R}^2$.

Ejemplo 6. *Considérese la función* $f(x,y,z) = x^2 + y^2 + z^2 - xy + yz - xz$ *definida en todo* $\mathbb{R}^3$. *Como* $\frac{\partial f}{\partial x}(x,y,z) = 2x - y - z$, $\frac{\partial f}{\partial y}(x,y,z) = -x + 2y + z$ *y* $\frac{\partial f}{\partial z}(x,y,z) = -x + y + 2z$ *para todo* $(x,y,z) \in \mathbb{R}^3$, *el único punto crítico de* f *es* $(x,y,z) = (0,0,0)$

y como $H_f(x,y,z) = H_f(0,0,0) = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}$ *para todo* $(x,y) \in \mathbb{R}^3$ *es una matriz definida positiva (aplíquese el criterio de Sylvester), en virtud del teorema 7, se sigue que* f *tiene un mínimo global estricto en* $(0,0,0)$.
Nótese que f *es diferenciable en todo* $\mathbb{R}^3$ *y por el teorema 5, el origen es el único punto crítico de* f*, luego* f *no tiene mas valores extremos.*

1.3. El método de los multiplicadores de Lagrange

En las secciones anteriores se han estudiado problemas de extremos libres tanto en funciones de una variable como en funciones de varias variables. En esta sección, se analizan problemas de *extremos condicionados* para funciones de varias variables mediante el clásico *método de los multiplicadores de Lagrange* y se muestra su eficacia con varios ejemplos de interés.

Como una breve introducción al *método de los multiplicadores de Lagrange*, considérese el siguiente problema de optimización. Supongamos que la función $f = f(x,y,z)$ representa la temperatura en el punto $(x,y,z) \in \mathbb{R}^3$ y nos preguntamos por el valor máximo y mínimo de la temperatura sobre una cierta superficie. Si la ecuación de la superficie está dada explícitamente como $z = h(x,y)$, entonces en la expresión $f(x,y,z)$ podemos sustituir z por $h(x,y)$ para obtener la temperatura sobre la superficie como una función $\tilde{f}(x,y) = f(x,y,h(x,y))$ que depende solamente de x e y. El problema se reduce entonces a encontrar los valores extremos de $\tilde{f}$ pero, en la práctica, surgen ciertas dificultades. La ecuación de la superficie podría darse en forma implícita como $g(x,y,z) = 0$ y ser imposible despejar z en función de x e y o igualmente x o y en función de las restantes variables. Incluso, el problema podría complicarse aún más al querer conocer los valores extremos de la temperatura sobre una determinada curva del espacio. Dicha curva será la intersección de dos superficies, digamos $g_1(x,y,z) = 0$ y $g_2(x,y,z) = 0$. Si pudiéramos despejar de estas dos ecuaciones x e y en función de z, introduciríamos estas expresiones en f y obtendríamos una nueva función con la única variable z, cuyos extremos podríamos entonces localizar con las técnicas descritas en las secciones previas. Sin embargo, esto no puede realizarse siempre y debe buscarse un método más efectivo. Un método muy elegante y útil para atacar este tipo de problemas fue desarrollado por el matemático italiano *Joseph-Louis Lagrange* (1736-1813).

Lagrange estudió derecho en el *College of Turín* (Turin, Italia) y, en un principio, no estaba muy fascinado en matemáticas. Su interés por esta ciencia se inició al leer la copia de un trabajo de E. Halley de 1693. De hecho, fue tanto su interés que Lagrange decidió estudiar matemáticas por su propia cuenta. El 23 de julio de 1754, Lagrange publicó su primer trabajo en matemáticas: una carta escrita en italiano y dirigida primero a su compatriota, el matemático G. Fagnano y, algo mas tarde, a L. Euler. En la carta, aparecía la analogía entre el binomio de Newton y las derivadas sucesivas del producto de funciones derivables. Sin embargo, al cabo de un mes, encontró su aparente descubrimiento en unas correspondencias entre Leibnitz y Bernoulli. Frustrado, se esforzó por hacer algo realmen-

te novedoso en matemáticas y comenzó a estudiar profundamente la curva tautócrona. A finales del año 1754, hizo importantes descubrimientos sobre esta curva, los cuales contribuyeron positivamente al desarrollo posterior de una nueva área en matemáticas: el *Cálculo de variaciones*.

El 12 de agosto de 1755, Euler recibió otra carta firmada por Lagrange, en la que mostraba una serie de resultados sobre la tautócrona y donde se podía ver también una primera aproximación a su método para el cálculo de máximos y mínimos condicionados, que más tarde sería nombrado en su honor: el *método de los multiplicadores de Lagrange*. A pesar de que Lagrange solo contaba con 19 años, el 28 de septiembre de ese mismo año ingresó como profesor de matemáticas en el *Royal Artillery School de Turín*. Al año siguiente, Lagrange vuelve a escribirle a Euler sobre los resultados que obtuvo al aplicar el Cálculo de variaciones a la Mecánica de la época. Euler quedó tan asombrado con su enorme talento que le propuso para las elecciones de la *Academia de Berlín* donde fue debidamente elegido el 2 de septiembre de 1756. A partir de entonces, comenzó una gran producción matemática por su parte.

Lagrange fue el mayor contribuyente en la revista científica *Mélanges de Turin* dividida en tres volúmenes (1759, 1762 y 1766). En dicha revista aparecen muchos de sus descubrimientos sobre el Cálculo de variaciones, una breve introducción al Cálculo de probabilidades, un extenso trabajo sobre los fundamentos de la Dinámica, un estudio sobre la propagación del sonido, contribuciones a la teoría de cuerdas, integración de ecuaciones diferenciales, Mecánica de fluidos y métodos para la resolución de sistemas lineales de ecuaciones diferenciales (los cuales usó para estudiar las órbitas de Júpiter y Saturno).

La *Academia de Ciencias de París*, le otorgó tres premios (1772, 1774 y 1780) gracias a sus contribuciones en el Problema de los tres cuerpos, el movimiento lunar y las órbitas de los planetas. En 1787, abandona Berlín y se traslada a Francia donde se convierte en miembro de la *Academia de Ciencias de París*. Al año siguiente, publicó su obra más importante, *Méchanique Analitique*, donde se resume toda la Mecánica conocida hasta el momento usando un fuerte desarrollo de la teoría de ecuaciones diferenciales y donde se aprecia ya de manera oficial y rigurosa su método para extremos condicionados (que tiene como origen la Mecánica y no la Matemática, como pudiera pensarse) [24]. En 1794, fundó la *École Polytechnique* donde impartía clases de Análisis matemático. En 1797 y 1800, Lagrange publica sus dos últimas obras: *Théorie des fonctions analytiques* (Teoría de funciones analíticas) y *Leçons sur le calcul des fonctions* (Lecciones sobre el cálculo de funciones), respectivamente.

En 1808, Napoleón le nombra *Legión de Honor* (la distinción francesa más importante del momento) y en 1813, fue galardonado con el premio *Grand Croix of the Ordre Impérial de la Réunion*. Desfortunadamente, falleció una semana después. Para más información sobre sus obras y biografía, el lector puede consultar [20].

Volviendo a cuestiones puramente matemáticas, se muestra a continuación la definición de *extremo condicionado* y se enuncia (sin demostración) el *Teorema de Lagrange*, que proporciona una condición necesaria para la existencia de extremos condicionados en

problemas de optimización bajo restricciones de igualdad.

Definición 9 (Extremo condicionado)**.** *Sea $n \in \mathbb{N}$ un número natural, $S \subseteq \mathbb{R}^n$ un subconjunto no vacío de $\mathbb{R}^n$, $C \subseteq S$ un subconjunto no vacío de S, $\boldsymbol{x_0} \in C$ un punto de C y $f : S \to \mathbb{R}$ una función real definida en S, entonces*

- *f tiene un **máximo global (condicionado)** por C en $\boldsymbol{x_0}$ si $f(\boldsymbol{x}) \leq f(\boldsymbol{x_0})$ para todo $\boldsymbol{x} \in C$.*
- *f tiene un **máximo local (condicionado)** por C en $\boldsymbol{x_0}$ si existe un entorno $N(\boldsymbol{x_0}) \subset S$ de $\boldsymbol{x_0}$ en S tal que $f(\boldsymbol{x}) \leq f(\boldsymbol{x_0})$ para todo $\boldsymbol{x} \in C \cap N(\boldsymbol{x_0})$.*
- *f tiene un **mínimo global (condicionado)** por C en $\boldsymbol{x_0}$ si $f(\boldsymbol{x}) \geq f(\boldsymbol{x_0})$ para todo $\boldsymbol{x} \in C$.*
- *f tiene un **mínimo local (condicionado)** por C en $\boldsymbol{x_0}$ si existe un entorno $N(\boldsymbol{x_0}) \subset S$ de $\boldsymbol{x_0}$ en S tal que $f(\boldsymbol{x}) \geq f(\boldsymbol{x_0})$ para todo $\boldsymbol{x} \in C \cap N(\boldsymbol{x_0})$.*

*La expresión **extremo (local o global) condicionado** se usa para referirse indistintamente a un máximo o mínimo (local o global) condicionado.*

Además, en la situación de la definición anterior, es claro que una función f tiene un máximo (resp. mínimo) local o global (condicionado) por C en un punto $\boldsymbol{x_0}$ si, y sólo si, $-f$ tiene un mínimo (resp. máximo) local o global (condicionado) por C en $\boldsymbol{x_0}$.

Teorema 8 (Teorema de Lagrange)**.** *Sean $n, m \in \mathbb{N}$ dos números naturales con $m < n$, $S \subseteq \mathbb{R}^n$ un subconjunto abierto y no vacío de $\mathbb{R}^n$, $f : S \to \mathbb{R}$ y $\boldsymbol{g} = (g_1, ..., g_m) : S \to \mathbb{R}^m$ dos funciones de clase $\mathcal{C}^1$ en S, X_0 el subconjunto de S dado por $X_0 = \{\boldsymbol{x} \in S : \ \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}\}$ y $\boldsymbol{x_0} \in X_0$ un punto de X_0. Supongamos que f tiene un extremo condicionado por X_0 en $\boldsymbol{x_0}$ y que la matriz jacobiana de $\boldsymbol{g} = (g_1, ..., g_m)$ en $\boldsymbol{x_0}$ tiene rango máximo m, entonces existen m números reales $\lambda_1, ..., \lambda_m \in \mathbb{R}$ verificando las siguientes n ecuaciones:*

$$D_i f(\boldsymbol{x_0}) + \sum_{j=1}^{m} \lambda_j D_i g_j(\boldsymbol{x_0}) = 0 \qquad (i = 1, ..., n) \tag{1.1}$$

La demostración tradicional del Teorema de Lagrange es algo compleja y hace uso del Teorema de la función implícita (razón por la cual se exige $m < n$ como hipótesis). No obstante, no vamos a reproducir la demostración de este teorema, ya que no es el objetivo del presente trabajo. El lector puede consultar dicha demostración en el capítulo 7 de [5]. En [8], se demuestra el Teorema de Lagrange en un ambiente de variedades diferenciables, donde intervienen los espacios tangente y normal a una variedad en un punto suyo. En el último capítulo de este trabajo, se prueba un resultado más general (teorema 9 y corolario 1) que incluye al Teorema de Lagrange como caso particular.

Nota 1. *En la situación del Teorema de Lagrange, las n ecuaciones (1.1) equivalen a la siguiente ecuación vectorial*

$$\nabla f(\boldsymbol{x_0}) + \lambda_1 \nabla g_1(\boldsymbol{x_0}) + \cdots + \lambda_m \nabla g_m(\boldsymbol{x_0}) = \boldsymbol{0} \tag{1.2}$$

Definición 10 (Condición y multiplicadores de Lagrange)**.** *En la situación del Teorema de Lagrange, la ecuación vectorial (1.2) asociada al problema de optimizar f condicionado por X_0 se denomina* ***condición de Lagrange*** *y los m números reales $\lambda_1, ..., \lambda_m$ reciben el nombre de* ***multiplicadores de Lagrange****.*

Es conveniente advertir que el recíproco del Teorema de Lagrange no es cierto. El siguiente ejemplo enseña que es posible que la condición de Lagrange tenga solución $\boldsymbol{x_0} \in \mathbb{R}^n$ y $\boldsymbol{\lambda} \in \mathbb{R}^m$ pero que $\boldsymbol{x_0}$ no sea ni máximo local ni mínimo local de una función f condicionado a restricciones $g_1, ..., g_m$.

Ejemplo 7 (¿Qué punto o puntos del elipsoide $E = \{(x, y, z) \in \mathbb{R}^3 : x^2 + \frac{y^2}{4} + \frac{z^2}{9} = 1\}$ están más próximos al orígen?)**.** *Basta minimizar la función distancia al cuadrado entre un punto $\boldsymbol{p} \in \mathbb{R}^3$ y el orígen, esto es, la función $d : \mathbb{R}^3 \to \mathbb{R}_0^+$ dada por $f(\boldsymbol{p}) = \|\boldsymbol{p}\|^2$ para todo $\boldsymbol{p} = (x, y, z) \in \mathbb{R}^3$ sujeta a la condición de que el punto pertenezca al elipsoide, esto es, que $g(x, y, z) = x^2 + \frac{y^2}{4} + \frac{z^2}{9} - 1 = 0$.*

$$\left.\begin{array}{r} \textit{minimizar } f(x, y, z) = x^2 + y^2 + z^2 \\ \textit{sujeto a } g(x, y, z) = x^2 + \frac{y^2}{4} + \frac{z^2}{9} - 1 = 0 \end{array}\right\}$$

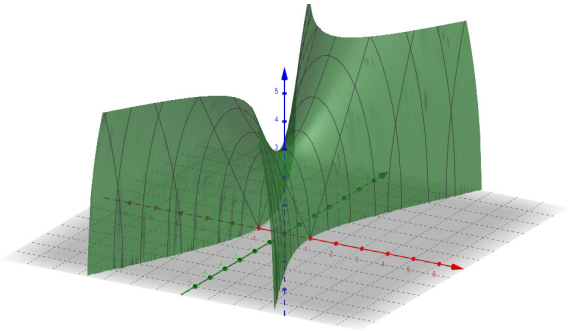


Figura 1.1: Elipsoide E

Las soluciones son $(1, 0, 0)$ y $(-1, 0, 0)$. Sin embargo, no es difícil comprobar que la condición de Lagrange asociada a este problema $\nabla f(x) + \lambda \nabla g(x) = 0$ se satisface precisamente en los puntos $(\pm 1, 0, 0)$, $(0, \pm 2, 0)$ y $(0, 0, \pm 3)$. De hecho, los puntos $(0, 0, \pm 3)$ son los máximos locales de f condicionados a $g(x, y, z) = 0$ para $\lambda = -9$, mientras que $(0, \pm 2, 0)$ son soluciones de la condición de Lagrange para $\lambda = -4$ que no son ni máximos ni mínimos locales de f condicionados a $g(x, y, z) = 0$.

A pesar de que el Teorema de Lagrange proporciona sólo una condición necesaria para optimizar una función sujeta a restricciones de igualdad, aporta un potente método para resolver tales problemas. A continuación, se resumen los pasos a seguir para aplicar correctamente el **método de los multiplicadores de Lagrange** a la hora de solucionar un tal problema de optimización condicionada:

En la situación del Teorema de Lagrange, considérese el problema de calcular los extremos de f condicionados por el conjunto $X_0 = \{\boldsymbol{x} \in S : \ \boldsymbol{g}(\boldsymbol{x}) = 0\}$, es decir, los valores

extremos de la expresión $f(x_1, ..., x_n)$ cuando las variables están sometidas a un cierto número de restricciones del tipo $g_1(x_1, ..., x_n) = 0, ..., g_m(x_1, ..., x_n) = 0$.

PASO 1: Construir la **función lagrangiana** $\mathcal{L} : S \times \mathbb{R}^m \longrightarrow \mathbb{R}$ asociada al problema. La función lagrangiana viene dada por la siguiente combinación lineal

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \langle \boldsymbol{\lambda}, g(\boldsymbol{x}) \rangle = f(\boldsymbol{x}) + \lambda_1 g_1(\boldsymbol{x}) + \cdots + \lambda_m g_m(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in S,\ \forall \boldsymbol{\lambda} \in \mathbb{R}^m \quad (1.3)$$

donde $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_m)$ y $\boldsymbol{g} = (g_1, ..., g_m)$.

PASO 2: Construir el **sistema de Lagrange** asociado a la correspondiente función lagrangiana. El sistema de Lagrange está definido como

$$\begin{cases} \nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = 0 \\ \boldsymbol{x} \in S \end{cases} \tag{1.4}$$

pero teniendo en cuenta que

$$\nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = (D_1 \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}), ..., D_n \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}), g_1(\boldsymbol{x}), ..., g_m(\boldsymbol{x}))$$

puede reescribirse el sistema anterior de la forma

$$\begin{cases} D_i f(\boldsymbol{x}) + \lambda_1 D_i g_1(\boldsymbol{x}) + \cdots + \lambda_m D_i g_m(\boldsymbol{x}) = 0 & (i = 1, ..., n) \\ g_j(\boldsymbol{x}) = 0 & (j = 1, ..., m) \\ \boldsymbol{x} \in S \end{cases} \tag{1.5}$$

o si se prefiere

$$\begin{cases} \nabla f(\boldsymbol{x}) + \lambda_1 \nabla g_1(\boldsymbol{x}) + \cdots + \lambda_m \nabla g_m(\boldsymbol{x}) = \boldsymbol{0} \\ g_j(\boldsymbol{x}) = 0 & (j = 1, ..., m) \\ \boldsymbol{x} \in S \end{cases} \tag{1.6}$$

que es un sistema (en general, no lineal) de $n + m$ ecuaciones con $n + m$ incógnitas (que son $\lambda_1, ..., \lambda_m$ y $x_1, ..., x_n$).

Lagrange descubrió que si el punto $(x_1, ..., x_n)$ es un extremo de f condicionado por X_0, entonces debe también cumplir el sistema de Lagrange asociado a f. Así pues, el Teorema de Lagrange puede leerse como: *los extremos de f condicionados por X_0 proceden de los puntos críticos de la función lagrangiana.*

PASO 3: Resolver el sistema de Lagrange (1.6) respecto a las incógnitas $x_1, ..., x_n$ y $\lambda_1, ..., \lambda_m$. Los multiplicadores de Lagrange $\lambda_1, ..., \lambda_m$, son introducidos únicamente como ayuda para resolver el sistema respecto a $x_1, ..., x_n$. Se introduce un multiplicador de Lagrange por cada restricción. Los puntos $(x_1, ..., x_n)$ así obtenidos deben entones comprobarse para determinar si originan un máximo condicionado, un mínimo condicionado o ninguno.

Respecto a esto último, existe un complicado criterio analítico para distinguir entre máximos y mínimos condicionados en tales problemas, esto es, una condición suficiente para problemas de extremos condicionados (véase [43]). No obstante, este criterio no es muy útil en la práctica y en algunos problemas particulares suele ser más fácil hacer uso de otros medios adecuados (por ejemplo, consideraciones físicas o geométricas) para hacer esta distinción. He aquí un primer ejemplo sencillo:

Ejemplo 8. *Optimizar la función $f(x,y) = 2xy$ definida en todo $(x,y) \in \mathbb{R}^2$ condicionada por $g(x,y) = x^2 + y^2 \leq 1$*

El Teorema de Weiertrass asegura la existencia de máximo y mínimo global de $f(x,y)$ sujeto a $g(x,y) \leq 1$ ya que f es continua y $K = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ es compacto. Para 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 el teorema 5 y las derivadas parciales de f en dicho punto deben anularse, dando lugar a un único punto candidato:

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

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

Función lagrangiana:

$$\mathcal{L} : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$$

$$\mathcal{L}(x,y,\lambda) = 2xy + \lambda(x^2 + y^2 - 1)$$

Sistema de Lagrange:

$$\begin{cases} 2y + 2\lambda x = 0 \\ 2x + 2\lambda y = 0 \\ x^2 + y^2 = 1 \end{cases} \Longrightarrow \lambda = \pm 1$$

Solución:

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

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

Solo falta comprobar los valores que toma f en cada uno de los puntos obtenidos:

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

Se concluye que el máximo de f en K vale 1 y se alcanza en los puntos $\left(\pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{2}}{2}\right)$ y el mínimo de f en K vale -1 y se alcanza en los puntos $\left(\pm\frac{\sqrt{2}}{2}, \mp\frac{\sqrt{2}}{2}\right)$.

Las aplicaciones del método de Lagrange son muy variadas y númerosas. A continuación, se exponen una serie de ejemplos donde se usa el método para obtener importantes resultados del Análisis matemático.

Ejemplo 9. *Constantes de equivalencia óptimas entre las normas $\|\cdot\|_1$ y $\|\cdot\|_p$ de $\mathbb{R}^3$ con $1 < p < \infty$.*

Dado $n \in \mathbb{N}$, todas las normas en $\mathbb{R}^n$ son equivalentes, es decir, existen constantes $\alpha, \beta \in \mathbb{R}^+$ tales que $\alpha\|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}\|_p \leq \beta\|\boldsymbol{x}\|_1$ para todo $\boldsymbol{x} \in \mathbb{R}^n$. El problema que se plantea es determinar la mayor constante $\alpha \in \mathbb{R}^+$ y la menor constante $\beta \in \mathbb{R}^+$ con $\alpha\|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}\|_p \leq \beta\|\boldsymbol{x}\|_1$ para todo $\boldsymbol{x} \in \mathbb{R}^3$ o, equivalentemente, $\alpha \leq \|\boldsymbol{x}\|_p \leq \beta$ para todo $\boldsymbol{x} \in \mathbb{R}^3$ con $\|\boldsymbol{x}\|_1 = 1$, esto es, calcular el máximo y mínimo global de la función $\|\cdot\|_p$ sobre la esfera unidad para la norma $\|\cdot\|_1$. Por razones de simetría, se puede restringir la atención al conjunto $K = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_1, x_2, x_3 \geq 0, \ x_1 + x_2 + x_3 = 1\}$ y por el crecimiento de la función $t \mapsto t^{1/p}$ de $\mathbb{R}_0^+$ en $\mathbb{R}$, puede plantearse el problema en los siguientes términos:

Determinar los extremos globales condicionados por K de la siguiente función

$$f : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \longrightarrow \mathbb{R}, \qquad f(x_1, x_2, x_3) = x_1^p + x_2^p + x_3^p, \quad \forall (x_1, x_2, x_3) \in (\mathbb{R}_0^+)^3$$

La continuidad de f y la compacidad de K aseguran la existencia de máximo y mínimo globales de f en K. Para resolver el problema se distinguen siete casos según la cantidad de variables x_1, x_2, x_3 que se anulan. Así pues, considérese siete subconjuntos $K_1, ..., K_7$ de K de manera que en cada uno de ellos se aplicará el método de los multiplicadores de Lagrange, obteniendo los candidatos a extremos condicionados de f que aporta cada uno de estos subconjuntos.

$$K_1 = \{(1, 0, 0)\}, \ K_2 = \{(0, 1, 0)\}, \ K_3 = \{(0, 0, 1)\}$$

$$K_4 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_1, x_2 > 0, \ x_3 = 0, \ x_1 + x_2 + x_3 = 1\}$$
$$K_5 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_1, x_3 > 0, \ x_2 = 0, \ x_1 + x_2 + x_3 = 1\}$$
$$K_6 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_2, x_3 > 0, \ x_1 = 0, \ x_1 + x_2 + x_3 = 1\}$$
$$K_7 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_1, x_2, x_3 > 0, \ x_1 + x_2 + x_3 = 1\}$$

Los conjuntos K_1, K_2 y K_3 son triviales al estar compuestos por un único punto en el que f toma el valor 1. Seguidamente, se analiza el subconjunto K_4 y se calculan los posibles extremos de la función f condicionados por K_4 mediante el método de los multiplicadores de Lagrange.

Función de lagrangiana:

$$\mathcal{L} : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}^2 \longrightarrow \mathbb{R}$$

$$\mathcal{L}(x_1, x_2, x_3, \lambda_1, \lambda_2) = x_1^p + x_2^p + x_3^p + \lambda_1(x_1 + x_2 + x_3 - 1) + \lambda_2 x_3$$

Sistema de Lagrange

$$\begin{cases} px_1^{p-1} + \lambda_1 = 0 \\ px_2^{p-1} + \lambda_1 = 0 \\ px_3^{p-1} + \lambda_1 + \lambda_2 = 0 \\ x_1 + x_2 + x_3 = 0 \\ x_3 = 0 \\ x_1, x_2 > 0 \end{cases}$$

Solución:

$$(x_1, x_2, x_3) = \left(\frac{1}{2}, \frac{1}{2}, 0\right) \qquad (\lambda_1, \lambda_2) = \left(\frac{-p}{2^{p-1}}, \frac{p}{2^{p-1}}\right)$$

Así, $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$ *es el único punto de* K_4 *en el que* f *puede tener un extremo global. Análogamente, el mismo procedimiento se sigue para los subconjuntos* K_5 *y* K_6*, que proporcionan los candidatos a extremos condicionados* $\left(\frac{1}{2}, 0, \frac{1}{2}\right)$ *y* $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$*, respectivamente.*

Finalmente, se calculan los posibles extremos de la función f *condicionados por* K_7*.*

Función lagrangiana:

$$\mathcal{L} : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R} \longrightarrow \mathbb{R}$$

$$\mathcal{L}(x_1, x_2, x_3, \lambda) = x_1^p + x_2^p + x_3^p + \lambda(x_1 + x_2 + x_3 - 1)$$

Sistema de Lagrange:

$$\begin{cases} px_1^{p-1} + \lambda = 0 \\ px_2^{p-1} + \lambda = 0 \\ px_3^{p-1} + \lambda = 0 \\ x_1 + x_2 + x_3 = 1 \\ x_1, x_2, x_3 > 0 \end{cases}$$

Solución:

$$\left.\begin{array}{c} x_k = \left(-\frac{\lambda}{p}\right)^{\frac{1}{p-1}} \qquad \forall k = 1, 2, 3 \\ x_1 + x_2 + x_3 = 1 \Rightarrow 3\left(-\frac{\lambda}{p}\right)^{\frac{1}{p-1}} = 1 \Rightarrow \lambda = -\frac{p}{3^{p-1}} \end{array}\right\} \Rightarrow \begin{cases} (x_1, x_2, x_3) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \\ \lambda = -\frac{p}{3^{p-1}} \end{cases}$$

Así, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ *es el único punto de* K_7 *en el que* f *puede tener un extremo global.*

Basta ya evaluar f *en cada uno de los puntos obtenidos para afirmar que* f *alcanza el máximo global en los puntos* $\{(1,0,0), (0,1,0), (0,0,1)\}$ *con un valor de 1 y el mínimo global en el punto* $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ *con un valor de* $\frac{1}{3^{p-1}}$*, por lo tanto se tiene*

$$\left(\frac{1}{3^{p-1}}\right)^{1/p} = \frac{1}{3^{1/p^*}} \leq \|\boldsymbol{x}\|_p \leq 1 \qquad \forall x \in \mathbb{R}^3,\ \|\boldsymbol{x}\|_1 = 1$$

donde $p^* \in (1, \infty)$ *es el exponente conjugado de* p*, esto es,* $\frac{1}{p} + \frac{1}{p^*} = 1$*. Las cotas obtenidas son inmejorables. En consecuencia, las constantes de equivalencia óptimas de las normas* $\|\cdot\|_1$ *y* $\|\cdot\|_p$ *en* $\mathbb{R}^3$ *son*

$$\frac{1}{3^{1/p^*}}\|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}\|_p \leq \|\boldsymbol{x}\|_1 \qquad \forall \boldsymbol{x} \in \mathbb{R}^3$$

La siguiente nota generaliza el resultado recién probado.

Nota 2. *Dados* $n \in \mathbb{N}$ *y* $p \in (1, +\infty)$*, las constantes de equivalencia óptimas de las normas* $\|\cdot\|_1$ *y* $\|\cdot\|_p$ *en* $\mathbb{R}^n$ *son:*

$$\frac{1}{n^{1/p^*}}\|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}\|_p \leq \|\boldsymbol{x}\|_1 \qquad \forall \boldsymbol{x} \in \mathbb{R}^n$$

donde $p^* \in (1, +\infty)$ *es el exponente conjugado de* p*.*

Ejemplo 10 (Desigualdad de Cauchy-Schwarz)**.** *Dado* $n \in \mathbb{N}$ *un número natural, la desigualdad de Cauchy-Schwarz en* $\mathbb{R}^n$ *establece lo siguiente*

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| = |\sum_{k=1}^{n} x_k y_k| \leq \left(\sum_{k=1}^{n} x_k^2\right)^{1/2} \left(\sum_{k=1}^{n} y_k^2\right)^{1/2} = \|\boldsymbol{x}\|\|\boldsymbol{y}\| \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \tag{1.7}$$

Si $\boldsymbol{x} = 0$ *o* $\boldsymbol{y} = 0$*, entonces la desigualdad (1.7) es trivialmente cierta. Supóngase que* $\boldsymbol{x}, \boldsymbol{y} \neq 0$ *y sean* $\boldsymbol{\alpha} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ *y* $\boldsymbol{\beta} = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$*. Si* $\langle \boldsymbol{\alpha}, \boldsymbol{\beta}\rangle \leq 1$ *entonces* $\langle \boldsymbol{x}, \boldsymbol{y}\rangle \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$ *y esto índica que es suficiente probar que* $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq 1$ *siempre y cuando* $\|\boldsymbol{x}\|^2 = \|\boldsymbol{y}\|^2 = 1$*.*

Considérese el siguiente programa

$$\begin{cases} \text{optimizar } f(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y}\rangle \\ \text{sujeto a } \begin{cases} g_1(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x}\|^2 - 1 = 0 \\ g_2(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{y}\|^2 - 1 = 0 \end{cases} \end{cases}$$

El Teorema de Weiertrass asegura la existencia de solución del problema y entonces el Teorema de Lagrange implica la existencia de los escalares $\lambda_1, \lambda_2 \in \mathbb{R}$ *tales que*

$$(\boldsymbol{y}, \boldsymbol{x}) + 2\lambda_1(\boldsymbol{x}, \boldsymbol{0}) + 2\lambda_2(\boldsymbol{0}, \boldsymbol{y}) = 0 \Leftrightarrow \begin{cases} \boldsymbol{y} = -2\lambda_1 \boldsymbol{x} \\ \boldsymbol{x} = -2\lambda_2 \boldsymbol{y} \end{cases} \Leftrightarrow \begin{cases} y_k = -2\lambda_1 x_k \\ x_k = -2\lambda_2 y_k \end{cases} \qquad \forall k = 1, ..., n$$

luego,

$$1 = \|\boldsymbol{y}\|^2 = \sum_{k=1}^{n} y_k^2 = \sum_{k=1}^{n} (-2\lambda_1 x_k)^2 = 4\lambda_1^2 \sum_{k=1}^{n} x_k^2 = 4\lambda_1^2 \Rightarrow \lambda_1 = \pm\frac{1}{2}$$

$$1 = \|\boldsymbol{x}\|^2 = \sum_{k=1}^{n} x_k^2 = \sum_{k=1}^{n} (-2\lambda_2 y_k)^2 = 4\lambda_2^2 \sum_{k=1}^{n} y_k^2 = 4\lambda_2^2 \Rightarrow \lambda_2 = \pm\frac{1}{2}$$

En ambos casos, se deduce que $\boldsymbol{y} = \pm\boldsymbol{x}$*, en cuyo caso* $f(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ *toma los valores 1 y -1, que serán el máximo y mínimo de* f *condicionados a* g_1 *y* g_2*, respectivamente. Por tanto,*

$$-1 \leq \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq 1 \Longleftrightarrow |\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq 1$$

como se quería.

Nota 3. *Como consecuencia de la desigualdad de Cauchy-Schwarz se tiene la continuidad del producto escalar en un espacio pre-hilbertiano. Además, esta desigualdad se deduce también de la desigualdad aritmético-geométrica (corolario 4), véase [44].*

El próximo ejemplo muestra inductivamente que una matriz simétrica $A \in \mathfrak{M}_n(\mathbb{R})$ es diagonalizable en $\mathbb{R}$.

Ejemplo 11 (Toda matriz simétrica real es diagonalizable en $\mathbb{R}$)**.** *Sea* $n \in \mathbb{N}$ *un número natural y* $A \in \mathfrak{M}_n(\mathbb{R})$ *una matriz simétrica y cuadrada de orden* n *con coeficientes reales, entonces* A *admite* n *vectores propios* $(\boldsymbol{x^{(1)}}, ..., \boldsymbol{x^{(n)}} \in \mathbb{R}^n : A\boldsymbol{x^{(k)}} = -\lambda_k \boldsymbol{x^{(k)}},\ k = 1, ..., n)$ *ortogonales* $(\langle \boldsymbol{x^{(i)}}, \boldsymbol{x^{(j)}} \rangle = 0,\ \forall i \neq j)$ *y unitarios* $(\|\boldsymbol{x^{(k)}}\| = 1,\ k = 1, ..., n)$*.*

Para conseguir el primer vector propio de A*, considerérese el siguiente problema de maximización condicionada:*

$$(P_1) \quad \begin{cases} \textit{maximizar } f(\boldsymbol{x}) = \langle \boldsymbol{x}, A\boldsymbol{x}^T \rangle \\ \textit{sujeto a } g_1(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 - 1 = 0 \end{cases}$$

donde $\boldsymbol{x} = (x_1, ..., x_n) \in \mathbb{R}^n$*. El Teorema de Weiertrass garantiza la existencia de máximo global* $\boldsymbol{x^{(1)}}$ *para el problema* (P_1) *y entonces el Teorema de Lagrange implica la existencia del escalar* $\lambda_1 \in \mathbb{R}$ *tal que*

$$\nabla f(\boldsymbol{x^{(1)}}) + \lambda_1 \nabla g_1(\boldsymbol{x^{(1)}}) = 0$$

o equivalentemente

$$2A\boldsymbol{x^{(1)}} + 2\lambda_1 \boldsymbol{x^{(1)}} = 0$$

pero esto conlleva que $A\boldsymbol{x^{(1)}} = -\lambda_1 \boldsymbol{x^{(1)}}$*, luego* $\boldsymbol{x^{(1)}}$ *es un vector propio (asociado al valor propio* $-\lambda_1$*) y unitario.*

Sea $m \in \mathbb{N}$ *un número natural con* $m < n$ *y supóngase que se tienen* m *vectores propios* $\boldsymbol{x^{(1)}}, ..., \boldsymbol{x^{(m)}}$ *ortogonales y unitarios asociados a los valores propios* $\lambda_1, ..., \lambda_m$*, respectivamente. Para conseguir el* $m+1$ *vector propio, considérese el siguiente problema de maximización condicionada*

$$(P_m) \quad \begin{cases} \textit{maximizar } f(\boldsymbol{x}) = \langle \boldsymbol{x}, A\boldsymbol{x}^T \rangle \\ \textit{sujeto a } \begin{cases} g_1(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 - 1 = 0 \\ g_2(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{x^{(1)}} \rangle = 0 \\ \vdots \\ g_{m+1}(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{x^{(m)}} \rangle = 0 \end{cases} \end{cases}$$

De nuevo, el Teorema de Weiertrass garantiza la existencia de máximo global $\boldsymbol{x}^{(m+1)}$ *para el problema* (P_m) *y entonces el Teorema de Lagrange implica la existencia de escalares* $\lambda_1, ..., \lambda_{m+1} \in \mathbb{R}$ *tales que*

$$\nabla f(\boldsymbol{x}^{(m+1)}) + \lambda_1 \nabla g_1(\boldsymbol{x}^{(m+1)}) + \cdots + \lambda_{m+1} \nabla g_{m+1}(\boldsymbol{x}^{(m+1)}) = 0$$

o equivalentemente

$$2A\boldsymbol{x}^{(m+1)} + 2\lambda_1 \boldsymbol{x}^{(m+1)} + \lambda_2 \boldsymbol{x}^{(1)} + \cdots + \lambda_{m+1} \boldsymbol{x}^{(m)} = 0$$

pero de esta ecuación y del hecho de que $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(m+1)}$ *son vectores unitarios y ortogonales entre sí, se sigue que*

$$\langle 2\boldsymbol{x}^{(k)}, A\boldsymbol{x}^{(m+1)} \rangle + \lambda_{k+1} = 0 \quad \forall k = 1, ..., m$$

y como $\boldsymbol{x}^{(k)}$ *es un vector propio de* A *para cada* $k = 1, ..., m$*, existe* $\mu_k \in \mathbb{R}$ *para cada* $k = 1, ..., m$ *tal que* $A\boldsymbol{x}^{(k)} = \mu_k \boldsymbol{x}^{(k)}$*, luego*

$$-\frac{\lambda_k}{2} = \langle \boldsymbol{x}^{(k)}, A\boldsymbol{x}^{(m+1)} \rangle = \langle A^T \boldsymbol{x}^{(k)}, \boldsymbol{x}^{(m+1)} \rangle = \langle A\boldsymbol{x}^{(k)}, \boldsymbol{x}^{(m+1)} \rangle = \mu_k \langle \boldsymbol{x}^{(k)}, \boldsymbol{x}^{(m+1)} \rangle = 0$$

de donde $\lambda_k = 0$ *para* $k = 2, ..., m+1$*, lo que implica*

$$2A\boldsymbol{x}^{(m+1)} + 2\lambda_1 \boldsymbol{x}^{(m+1)} = 0$$

así que $\boldsymbol{x}^{(m+1)}$ *es un vector propio de* A*. Como* $\boldsymbol{x}^{(m+1)}$ *es unitario y* $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(m+1)}$ *son ortogonales entre sí, la prueba queda completada.*

Nota 4. *El resultado probado en el ejemplo anterior es conocido como Teorema espectral para matrices simétricas reales. Obsérvese que*

(i) El conjunto de puntos $\boldsymbol{x} \in \mathbb{R}^n$ *tales que* $g_1(\boldsymbol{x}) = 0, ..., g_k(\boldsymbol{x}) = 0$ *es no vacío hasta* $k = n-1$*, así que el proceso inductivo llevado a cabo en el ejemplo anterior se detiene al llegar a* $\boldsymbol{x}^{(n)}$*, como era de esperar.*

(ii) No se ha comprobado explícitamente la hipótesis de que $\{\nabla g_1(\boldsymbol{x}^{(1)}), ..., \nabla g_m(\boldsymbol{x}^{(m+1)})\}$ *sea un conjunto linealmente independiente, pero efectivamente esto sucede puesto que* $\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(m+1)}\}$ *son vectores ortogonales entre si.*

Ejemplo 12 (Desigualdad de Hadamard)**.** *Sea* $X = (x_{ij}) \in \mathfrak{M}_m(\mathbb{R})$ *una matriz de orden* $m \in \mathbb{N}$ *con coeficientes reales y* $\boldsymbol{x}^{(k)} = (x_{k1}, ..., x_{km})$ *la fila* k*-ésima de* X *para* $k = 1, ..., m$*, entonces*

$$|\det(X)| \leq \prod_{k=1}^{m} \|\boldsymbol{x}^{(k)}\|$$

donde $\|\boldsymbol{x}^{(k)}\|^2 = x_{k1}^2 + \cdots + x_{km}^2 = \theta_m^2$ *para* $k = 1, ..., m$*.*

Considérese el programa

$$\begin{cases} \textit{maximizar } f(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) = \det(X) \\ \textit{sujeto a } \begin{cases} g_1(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) = \|\boldsymbol{x^{(1)}}\|^2 - \theta_1^2 = 0 \\ \vdots \\ g_n(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) = \|\boldsymbol{x^{(m)}}\|^2 - \theta_m^2 = 0 \end{cases} \end{cases}$$

El Teorema de Weiertrass asegura la existencia de solución del problema y entonces el Teorema de Lagrange implica la existencia de multiplicadores $\lambda_1,...,\lambda_m \in \mathbb{R}$ *no todos nulos (se puede tomar* $\lambda_0 = 1$*) tales que*

$$\nabla f(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) + \sum_{k=1}^{m} \lambda_k \nabla g_k(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) = 0$$

El desarrollo de Laplace del determinante de una matriz asegura

$$\nabla f(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) = (\Delta_{11},...,(-1)^{1+m}\Delta_{1m},...,(-1)^{m+1}\Delta_{m1},...,\Delta_{mm})$$

donde Δ_{ij} *representa el determinante de la matriz de orden* $m-1$ *obtenida al eliminar la fila* i*-ésima y la columna* j*-ésima de* X*. Además,*

$$\nabla g_1(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) = 2(x_{11},...,x_{1m},0,...,0)$$

$$\vdots$$

$$\nabla g_k(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) = 2(0,...,0,x_{k1},...,x_{km},0,...,0)$$

$$\vdots$$

$$\nabla g_m(x_{11},...,x_{1m},...,x_{m1},...,x_{mm}) = 2(0,...,0,x_{m1},...,x_{mm})$$

luego la condición de Lagrange es

$$\left.\begin{array}{c} \Delta_{11} + 2\lambda_1 x_{11} = 0 \\ \vdots \\ (-1)^{1+m}\Delta_{1m} + 2\lambda_1 x_{1m} = 0 \end{array}\right\} \Longrightarrow x_{11} = -\frac{\Delta_{11}}{2\lambda_1},...,x_{1m} = (-1)^m \frac{\Delta_{1m}}{2\lambda_1}$$

$$\vdots$$

$$\left.\begin{array}{c} (-1)^{m+1}\Delta_{m1} + 2\lambda_m x_{m1} = 0 \\ \vdots \\ \Delta_{mm} + 2\lambda_m x_{mm} = 0 \end{array}\right\} \Longrightarrow x_{m1} = (-1)^m \frac{\Delta_{m1}}{2\lambda_m},...,x_{mm} = -\frac{\Delta_{mm}}{2\lambda_m}$$

(obviamos la posibilidad $\lambda_k = 0,\ k = 1,...,m$ *pues en ese caso,* $\det(X) = 0$ *y la desigualdad de Hadamard es trivial). Considérese ahora la matriz* $(y_{ij}) = Y = XX^T$ *cuyos coeficientes son los siguientes*

$$y_{ij} = \sum_{k=1}^{m} x_{ik}x_{jk} = -\frac{1}{2\lambda_i}\sum_{k=1}^{m}(-1)^{i+k}\ \Delta_{ik}\ x_{jk} = \begin{cases} \theta_i^2 & \textit{si } i = j \\ 0 & \textit{si } i \neq j \end{cases}$$

Por tanto,

$$Y = XX^T = \begin{pmatrix} \theta_1^2 & & \\ & \ddots & \\ & & \theta_m^2 \end{pmatrix} \Longrightarrow \det(Y) = \det(X)^2 = \prod_{k=1}^{m} \theta_k^2$$

Esto es, en cualquier punto crítico donde $\det(X) \neq 0$*, se tiene que* $\det(X)^2 = \prod_{k=1}^{m} \|\boldsymbol{x}^{(k)}\|^2$*, es decir*

$$|\det(X)| = \pm \prod_{k=1}^{m} \|\boldsymbol{x}^{(k)}\|$$

Obviamente, el signo + corresponde al máximo global y el signo -, al mínimo global y se concluye que

$$|\det(X)| \leq \prod_{k=1}^{m} \|\boldsymbol{x}^{(k)}\|$$

como se quería.

Nota 5. *Esta desigualdad fue demostrada por primera vez en 1893 por el matemático francés Jacques Hadamard. La desigualdad también se cumple para columnas pues el determinante de una matriz coincide con el de la matriz transpuesta.*

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 ocasiones, se suelen 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 capítulo 3 de [40]. En otras ocasiones, se suelen utilizar *artificios* especiales a fin de conseguir los valores extremos de una función directamente sin hallar de antemano los puntos particulares en los que tales extremos se alcanzan. El siguiente ejemplo pone de manifiesto uno de estos artificios.

Ejemplo 13. *Sean* $A, B, C, D, E, F \in \mathbb{R}$ *constantes y consideremos la superficie cuádrica con centro en el origen dada por la ecuación* $Ax^2 + By^2 + Cz^2 + 2Dyz + 2Exz + 2Fxy = 1$*. El problema consiste en determinar las longitudes de los semiejes.*

Escribimos $\boldsymbol{x} = (x_1, x_2, x_3)$ *en vez de* (x, y, z) *para acomodar la notación e introducimos la forma cuadrática*

$$q(x) = \sum_{j=1}^{3} \sum_{i=1}^{3} \alpha_{ij} x_i x_j \qquad \forall \boldsymbol{x} \in \mathbb{R}^3$$

donde elegimos los coeficientes $\alpha_{ij} = \alpha_{ji}$ *de forma que la ecuación de la cuádrica sea* $q(x) = 1$ *(luego la forma cuadrática es simétrica y definida positiva). El problema equivale a optimizar la función* $f(\boldsymbol{x}) = x_1^2 + x_2^2 + x_3^2$ *sujeta a la condición* $g(\boldsymbol{x}) = 0$ *con* $g(\boldsymbol{x}) = q(\boldsymbol{x}) - 1$*.*

Utilizamos el método de los multiplicadores de Lagrange e introducimos un multiplicador. Como $\nabla g = \nabla q$*, se considera la ecuación vectorial* $\nabla f(\boldsymbol{x}) + \lambda \nabla q(\boldsymbol{x}) = 0$*. En este caso, tanto* f *como* g *son funciones homogéneas de grado dos,*

$$f(\lambda \boldsymbol{x}) = \lambda^2 f(\boldsymbol{x}),\ g(\lambda \boldsymbol{x}) = \lambda^2 g(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \mathbb{R}^3, \forall \lambda \in \mathbb{R}$$

a las que se les puede aplicar el Teorema de Euler para funciones homogéneas

$$\langle \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle = 2f(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \mathbb{R}^3$$

y, así, la ecuación vectorial anterior queda

$$\langle \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle + \langle \lambda \boldsymbol{x}, \nabla q(\boldsymbol{x}) \rangle = 2f(\boldsymbol{x}) + 2\lambda q(\boldsymbol{x}) = 0$$

Puesto que $q(\boldsymbol{x}) = 1$ *sobre la superficie, determinamos* $\lambda = -f(\boldsymbol{x})$ *y llamando* $t = 1/f(\boldsymbol{x})$ *(obsérvese que* $f(\boldsymbol{x}) \neq 0$ *en la superficie), la ecuación* $\nabla f(\boldsymbol{x}) + \lambda \nabla q(\boldsymbol{x}) = 0$ *se convierte en* $t\nabla f(\boldsymbol{x}) - \nabla q(\boldsymbol{x}) = 0$*. La ecuación vectorial* $t\nabla f(\boldsymbol{x}) - \nabla q(\boldsymbol{x}) = 0$ *conduce a tres ecuaciones para* x_1, x_2 *y* x_3*:*

$$\begin{array}{ccccccc} (a_{11}-t)x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = & 0 \\ a_{21}x_1 & + & (a_{22}-t)x_2 & + & a_{23}x_3 & = & 0 \\ a_{31}x_1 & + & a_{32}x_2 & + & (a_{33}-t)x_3 & = & 0 \end{array}$$

Como este sistema es homogéneo y $\boldsymbol{x} = \mathbf{0}$ *no es solución del problema, el determinante de la matriz correspondiente al sistema debe anularse:*

$$\begin{vmatrix} a_{11}-t & a_{12} & a_{13} \\ a_{21} & a_{22}-t & a_{23} \\ a_{31} & a_{32} & a_{33}-t \end{vmatrix} = 0$$

Esta ecuación se conoce como ecuación característica de la forma cuadrática q*. En este caso, la naturaleza geométrica del problema garantiza que las tres raíces* t_1, t_2 *y* t_3 *deben ser reales y positivas (debido a que* q *es simétrica y definida positiva).*

Los semiejes de la superficie cuádrica son entonces $t_1^{-1/2}, t_2^{-1/2}$ *y* $t_3^{-1/2}$*.*

Aunque todos los ejemplos aquí expuestos son puramente matemáticos, no hay que olvidar que el Teorema de Lagrange tiene su orígen en la Mecánica y como tal, también goza de numerosas aplicaciones en Física, Ingeniería o incluso Economía, entre otras. Véase, por ejemplo, [23] y [39].

Capítulo 2

Orígenes históricos de la Programación no lineal

El *Teorema de Kuhn-Tucker* es el primer y principal resultado de toda una teoría que se desarrollo a su alrededor y que dio lugar a la *Programación no lineal*. Así que analizando el origen de este teorema, quedará de manifiesto el origen de la Programación no lineal.

Cuando Kuhn y Tucker demostraron su teorema lanzaron la teoría de la Programación no lineal y la posterior investigación en esta rama de la matemática. Sin embargo, en cierto modo este teorema ya había sido demostrado antes: en 1939 por W. Karush en su tesis de fin de máster que no fue publicada en su día, en 1948 por F. John en un artículo que fue rechazado por el *Duke Mathematical Journal* y, posiblemente algo antes por Ostrogradsky y Farkas. En este capítulo se analizan los resultados que probaron cada uno de estos matemáticos teniendo en cuenta el campo de la matemática a la que pertenecían y donde se consideraron una contribución. En el camino se discutirán preguntas como si el Teorema de Kuhn y Tucker debe ser considerado un descubrimiento múltiple y porqué las tres apariciones del mencionado teorema supusieron reacciones tan distintas por parte de la comunidad matemática del momento. La importancia del contexto social y matemático estará presente, así como la influencia de la *Investigación operativa* (IO) y el papel jugado por la *Oficina de Investigación Naval* (OIN).

2.1. Introducción

En el verano de 1950, en el *Segundo Simposio de Berkeley* (California) sobre matemática estadística y probabilidad, un reconocido topólogo de Princeton, Albert W. Tucker, dió una conferencia bajo el título *Programación no lineal*, basada en un trabajo conjunto con Harold. W. Kuhn, quien acababa de terminar su tesis doctoral en la Universidad de Princeton. El trabajo de Kuhn y Tucker fue publicado tras la conferencia con el nombre de *Programación no lineal*. Esta fue la primera vez que el nombre de "programación no lineal" aparecía en la literatura matemática. En el trabajo, Kuhn y Tucker introducen un problema de Programación no lineal y prueban el principal teorema de esta teoría (el conocido Teorema de Kuhn-Tucker), que proporciona condiciones necesarias para la existencia de una solución óptima para un problema de Programación no lineal.

El resultado se popularizó casi de inmediato y, no mucho más tarde de su publicación, la gente empezó a referirse a él como el *Teorema de Kuhn-Tucker*, pero aparentemente Kuhn y Tucker no fueron los primeros matemáticos en demostrarlo. William Karush demostró el mismo teorema en 1939 en su tesis de fin de máster por la Universidad de Chicago y Fritz John obtuvo (casi) el mismo resultado en un artículo publicado en 1948 para una recopilación de tratados con motivo del 60º cumpleaños del matemático alemán Richard Courant. En la actualidad, se suele referir al teorema como el *Teorema de Karush-Kuhn-Tucker*, para así reconocer el trabajo de Karush, a pesar de que cuando él entregó su tesis de fin de máster en diciembre de 1939, ésta no fue publicada, nadie le animó a hacerlo y, en general, no parecía ser muy interesante o atractiva para los matemáticos de aquella época. El artículo de F. John fue escrito sólo dos años antes que el de Kuhn y Tucker. John intentó publicarlo en el *Duke Mathematics Journal* pero rechazaron su artículo. Es chocante que sólo dos años después cuando surge el Teorema de Kuhn y Tucker, éste fuera reconocido, adquiriendo fama enseguida y dando lugar al posterior desarrollo de una nueva área de investigación matemática.

Obviamente, esto hace que nos planteemos las siguientes preguntas: ¿Realmente eran el mismo resultado? ¿Es justo hablar de un descubrimiento múltiple? ¿Por qué la comunidad matemática reaccionó de manera tan diferente en los tres casos? ¿Por qué no pasó absolutamente nada en los dos primeros casos? O mejor dicho, ¿Por qué el Teorema de Kuhn y Tucker tuvo ese impacto tan grande en la comunidad matemática del momento?. Antes de seguir con el capítulo y responder estas preguntas será conveniente resaltar algunos prerrequisitos matemáticos.

2.1.1. Prerrequisitos matemáticos

A continuación se explica de forma breve lo que se entiende por un problema de programación no lineal o un *programa no lineal* y se enuncia de manera precisa el mencionado Teorema de Kuhn-Tucker.

Dados $n, m \in \mathbb{N}$, un problema de Programación no lineal es un problema de optimización finito-dimensional consistente en minimizar una determinada función (función objetivo) $f : \Omega \to \mathbb{R}$ definida en un subconjunto Ω de $\mathbb{R}^n$, siendo $g_1, ..., g_m$ funciones reales definidas en Ω:

$$\left.\begin{array}{c} \text{minimizar } f(\boldsymbol{x}) \text{ sujeto a} \\ g_1(\boldsymbol{x}) \leq 0, ..., g_m(\boldsymbol{x}) \leq 0 \\ \boldsymbol{x} = (x_1, ..., x_n) \in \Omega \end{array}\right\} \qquad (P)$$

De manera análoga se puede considerar el problema de maximizar la función f. Como maximizar una función f es equivalente a minimizar la función $-f$ vamos a restringirnos únicamente al caso de minimizar. Enunciamos ya el Teorema de Kuhn-Tucker, que estudiaremos de forma precisa en el último capítulo del trabajo, pero de momento, basta con enunciarlo simplemente para tener una idea de sobre que trata.

Teorema de Kuhn-Tucker. Dados $n, m \in \mathbb{N}$ dos números naturales, $\Omega \subset \mathbb{R}^n$ un subconjunto abierto y no vacío de $\mathbb{R}^n$ y $\overline{\boldsymbol{x}} \in \Omega$ un punto satisfaciendo todas las restricciones de (P) y cumpliendo que los vectores gradientes $\nabla g_k(\overline{\boldsymbol{x}})$ de las restricciones con $g_k(\overline{\boldsymbol{x}}) = 0$ son linealmente independientes para todo $k = 1, ..., m$. Supongamos que las funciones $f, g_1, ..., g_m$ son diferenciables en $\overline{\boldsymbol{x}}$ y que $\overline{\boldsymbol{x}}$ es un mínimo local para el problema (P), entonces existen escalares $u_1, ..., u_m \in \mathbb{R}$ (multiplicadores) tales que

$$\nabla f(\overline{\boldsymbol{x}}) + \sum_{k=1}^{m} u_k \nabla g_k(\overline{\boldsymbol{x}}) = 0 \tag{2.1}$$

$$u_k g_k(\overline{\boldsymbol{x}}) = 0 \quad \forall k = 1, ..., m \tag{2.2}$$

$$u_k \geq 0 \quad \forall k = 1, ..., m \tag{2.3}$$

El conjunto de estas tres condiciones necesarias para la existencia de mínimo local para (P) reciben el nombre de *condiciones de Kuhn-Tucker*. La condición (2.1) establece que la correspondiente función lagrangiana $\phi(\overline{\boldsymbol{x}}, \boldsymbol{u}) = f(\overline{\boldsymbol{x}}) + \sum_{k=1}^{m} u_k g_k(\overline{\boldsymbol{x}})$ tiene un punto crítico en $(\overline{\boldsymbol{x}}, \boldsymbol{u})$. La condición (2.2) establece que si $g_k(\overline{\boldsymbol{x}}) \neq 0$ entonces $u_k = 0$. La condición (2.3) establece que los multiplicadores han de ser no negativos.

En lo que sigue, se analizan ya los tres orígenes tan diversos del Teorema de Kuhn-Tucker, comenzando por el resultado obtenido por W. Karush, seguido del resultado de F. John y terminando con el propio de Kuhn y Tucker.

2.2. El Teorema de Karush y el Cálculo de variaciones

En diciembre de 1939, William Karush (1917-1997) terminó sus estudios de máster en Matemáticas por la Universidad de Chicago. El título de su tesis de fin de máster fue *Minima of Functions of Several Variables with Inequilities as Side Conditions* [25]. Actualmente, diríamos que tal problema de optimización pertenece al campo de la Programación no lineal, pero esta área no existía aún en aquellos tiempos. Así que, debemos profundizar más en la tesis de Karush y el ambiente que le rodeaba para lograr determinar el campo de la matemática a la que pertenecía y en el cual se consideró una contribución. Dicha tesis fue supervisada por el profesor Lawrence M. Graves y en la introducción de su tesis, Karush señala la motivación de su trabajo escribiendo lo siguiente:

> *El problema de determinar condiciones necesarias y suficientes para la existencia de un mínimo local de una función $f = f(x_1, ..., x_n)$ sobre el conjunto de puntos $\boldsymbol{x} = (x_1, ..., x_n)$ que verifican las ecuaciones $g_\alpha(\boldsymbol{x}) = 0$ con $\alpha = 1, ..., m$ y donde las funciones f y g_α $(\alpha = 1, ..., m)$ tienen derivadas continuas de, al menos, segundo orden, ha sido satisfactoriamente tratado [31]. Este artículo [la tesis de Karush] se propone extender el correspondiente problema a la clase de puntos $\boldsymbol{x}$ verificando las desigualdades $g_\alpha(\boldsymbol{x}) \geq 0$ para $\alpha = 1, ..., m$ donde m puede ser menor, igual o mayor que n*

El problema que Graves le propuso a Karush para su tesis surge de un intento de extender el trabajo sobre Cálculo de variaciones de Gilbert A. Bliss [31], quien fue jefe del departamento de matemáticas en la Universidad de Chicago, a un caso más general. Consecuentemente, las raíces del problema que Karush presentó en su tesis se escondían en el Cálculo de variaciones, un campo de la matemática íntimamente relacionado con el departamento de matemáticas en Chicago.

En aquella época, el departamento de matemáticas de la Universidad de Chicago, fundado con la apertura de la misma en 1892, estaba dirigido por E. H. Moore (1862-1932) junto con G. O. Bolza (1857-1936) y H. Maschke (1853-1903), quienes crearon un ambiente matemático que condujo al departamento a ser uno de los mas influyentes en matemáticas en USA. Fue, precisamente, Bolza quien se interesó obsesionadamente por el Cálculo de variaciones, llegando a crear un grupo sólido y amplio de investigación en este campo, conocido como *La Escuela de Chicago en cálculo de variaciones*, o simplemente, la *Escuela de Chicago*. Bolza fue un buen director de tesis y a menudo guiaba a sus alumnos para que acabasen trabajando en la rama de las matemáticas en la que él investigaba. La Escuela de Chicago fue el resultado de numerosas tutorías y seminarios por parte de Bolza con sus propios alumnos sobre temas de Cálculo de variaciones. La mayoría de estos estudiantes acabaron realizando sus tesis bajo la supervisión de Bolza que, como era de esperar, trataba sobre el Cálculo de variaciones, donde Bolza era considerado un gran investigador. Bolza se interesó en este campo a través de una conferencia de Weiertrass en 1879.

En 1908, Maschke falleció y dos años después Bolza regresó a Alemania, su país natal. Chicago perdió, así, dos de sus líderes matemáticos, lo que derivó en un declive en la reputación del departamento que supuso la llegada a éste de un "nuevo equipo" formado por Bliss, Dickson y Wliczynski. Entre 1927 y 1941, el nuevo departamento y sobretodo Bliss, que fue alumno de Bolza, continuaron con la tradición de los anteriores líderes y se caracterizó por un intenso estudio en Cálculo de variaciones que ocupó la mayor parte de la investigación matemática en Chicago. De hecho, en el periodo de 10 años comprendido entre 1927 y 1937, Bliss dirigió 35 tesis doctorales, de las cuales 34 pertenecían al Cálculo de variaciones.

Como estudiante en Chicago, Karush fue producto de esta tradición y su tesis de máster debe ser analizada y discutida en este contexto. El objetivo de su tesis era determinar condiciones necesarias y suficientes para la existencia de mínimo local de una función $f(x_1, ..., x_n)$ en el conjunto de puntos $(x_1, ..., x_n)$ que satisfacen las desigualdades $g_1(\boldsymbol{x}) \geq 0, ..., g_m(\boldsymbol{x}) \geq 0$, donde se exigía una cierta regularidad a las funciones involucradas f y $g_1, ..., g_m$. Karush llevó a cabo su trabajo en 1939 mientras que la Escuela de Chicago se centraba en problemas de Cálculo variacional con restricciones de desigualdad, esto es, problemas del tipo minimizar o maximizar el funcional

$$\mathcal{F}[\varphi] = \int_I F(\boldsymbol{x}, \varphi(\boldsymbol{x}), \varphi'(\boldsymbol{x}))\ dx$$

en el conjunto $\mathcal{D} = \{\varphi \in \mathcal{C}^1(I) : \varphi(a) = A, \varphi(b) = B\}$, donde $I = [a, b] \subset \mathbb{R}$ es un intervalo cerrado no trivial, $D \subset \mathbb{R}^2$ es un dominio de $\mathbb{R}^2$ y $F : I \times D \to \mathbb{R}$ es una función

conocida.

Así pues, el trabajo de Karush se concibió como una versión finito-dimensional de tales problemas y que, por consiguiente, carecía de interés. Sin embargo, Karush no interpretó su trabajo como un caso finito-dimensional de los problemas infinito-dimensionales que se trataban en el Cálculo de variaciones y que realmente eran los que llamaban la atención, sino como una extensión del artículo de Bliss, publicado el año anterior. Desde mediados de los años 30, Bliss se había interesado en unas propiedades llamadas *normalidad* y *anormalidad* para la curva que minimiza una cierta función condicionada a restricciones de igualdad. El objetivo del artículo de Bliss, el cual Karush usó como punto de partida en su tesis, era como dijo el propio Bliss *"analizar mas específicamente de lo hecho hasta entonces el significado de normalidad y anormalidad en el Cálculo de variaciones. Para hacer esto, hago incapié en el artículo [31] en el significado de normalidad y anormalidad para el problema del mínimo local de una función de un finito número de variables"*.

El resultado de Karush, relacionado con el Teorema de Kuhn-Tucker, se enuncia y se demuestra en la tercera parte de su tesis. En ella, Karush estudia el mínimo de una función $f(\boldsymbol{x})$ sujeto a las condiciones $g_1(\boldsymbol{x}) \geq 0, ..., g_m(\boldsymbol{x}) \geq 0$, donde tanto f como $g_1, ..., g_m$ son de clase $\mathcal{C}^1$ en un entorno de un punto $\boldsymbol{x^0} \in \Omega$. Pero antes de demostrar el Teorema de Kuhn-Tucker, Karush mostró una versión de éste menos restrictiva:

Si $f(\boldsymbol{x^0})$ es un mínimo, entonces existen multiplicadores l_0 y l_α no todos cero tal que las derivadas $\frac{\partial F}{\partial x_1}, ..., \frac{\partial F}{\partial x_n}$ de la función $F(\boldsymbol{x}) = l_0 f(\boldsymbol{x}) + l_\alpha g_\alpha(\boldsymbol{x})$ se anulan en $\boldsymbol{x^0}$

Hay que mencionar que aquí, Karush utiliza el *criterio de sumación de Einstein*, es decir, $l_\alpha g_\alpha(\boldsymbol{x})$ significa $\sum_{\alpha=1}^{m} l_\alpha g_\alpha(\boldsymbol{x})$.

Nótese que no hay restricción para el signo de los multiplicadores. Además, el multiplicador l_0, asociado a la función objetivo, puede tomar el valor 0. En este caso, Karush denominaba al punto $\boldsymbol{x^0}$, *punto anormal*. Para evitar el caso anormal, se necesitaban asumir algunas condiciones adicionales sobre la función f y las funciones $g_1, ..., g_m$ (la *constraint qualification* como mas tarde la denominaron Kuhn y Tucker). Karush estableció tales condiciones, introduciendo los conceptos de *dirección admisible*, *curva admisible* y *punto normal*.

Por dirección admisible, Karush entendía un vector no nulo $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)$ que resuelve el sistema de desigualdades

$$\sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i}(\boldsymbol{x^0})\lambda_i \geq 0 \qquad \forall j = 1, ..., m$$

En otras palabras, él consideraba al vector $\boldsymbol{\lambda} \neq \boldsymbol{0}$ una dirección admisible si la derivadas direccionales de las funciones $g_1, ..., g_m$ en la dirección de $\boldsymbol{\lambda}$ eran no negativas, lo que significa que "uno permanece" en la zona factible si "uno camina" desde $\boldsymbol{x^0}$ en la dirección de $\boldsymbol{\lambda}$.

Fijado $t_0 > 0$, por curva admisible, Karush entendía un arco regular $x_i(t)$ $(i = 1, ..., n,\ 0 \leq t \leq t_0)$ verificando $g_j(x_i(t)) \geq 0$ para todo $j = 1, ..., m$ y todo $t \in [0, t_0]$. Lo que significa que "uno permanece" en la zona factible cuando "uno se mueve" a lo largo del arco.

Por punto normal, Karush entendía un punto $\boldsymbol{x^0}$ satisfaciendo que la matriz jacobiana de $\boldsymbol{g} = (g_1, ..., g_m)$ tiene rango m o, lo que es lo mismo, los vectores gradientes $\nabla g_1(\boldsymbol{x^0}), ..., \nabla g_m(\boldsymbol{x^0})$ son linealmente independientes.

Llegado a este punto, Karush formuló el Teorema de Kuhn-Tucker de la siguiente manera:

> *Supóngase que para cada dirección admisible $\boldsymbol{\lambda}$, existe una curva admisible partiendo de $\boldsymbol{x^0}$ en la dirección de $\boldsymbol{\lambda}$. Entonces una primera condición necesaria para que $f(\boldsymbol{x^0})$ sea un mínimo de f es que existan multiplicadores $l_\alpha \leq 0$ tales que todas las derivadas $\frac{\partial F}{\partial x_1}, ..., \frac{\partial F}{\partial x_n}$ de la función $F = f + l_\alpha g_\alpha$ se anulen en $\boldsymbol{x^0}$*

donde por un arco $x_i(t)$ partiendo de $\boldsymbol{x^0}$ en la dirección $\boldsymbol{\lambda}$, entendía que $x_i(0) = x_i^0$ y $x_i'(0) = \lambda_i$ para $i = 1, ..., n$.

Su idea era usar el Lema de Farkas (véase [42]) para garantizar la existencia de multiplicadores no positivos. Ciertamente, el Teorema de Karush mencionado arriba se parece bastante al teorema original de Kuhn-Tucker: deben existir multiplicadores (l_α) tales que la función F tiene un punto crítico en $\boldsymbol{x^0}$. La condición $l_\alpha g_\alpha(\boldsymbol{x^0}) = 0$ no aparece porque Karush solo consideraba restricciones con $g_\alpha(\boldsymbol{x^0}) = 0$.

En 1975, Kuhn escribió una carta dirigida a Karush, reconociendo la prioridad de éste sobre el resultado conocido por Teorema de Kuhn-Tucker. Con esto, queda claro que Kuhn era consciente del trabajo de Karush. De hecho, en una publicación suya posterior menciona la tesis de Karush como un artículo clásico en el campo de la Programación no lineal que no fue publicado hasta la fecha. Esta publicación de Kuhn, pone de manifiesto la interpretación de la tesis de Karush como un artículo propio del campo de la Programación no lineal y no del Cálculo de variaciones como fue visto en un principio y que provocó que no se le diera la correspondiente importancia ni, mucho menos, que fuese publicado.

Por otra parte, la reacción de Tucker ante esta situación fue similar a la de Kuhn. Tucker quedó sorprendido ya que tuvo la oportunidad de conocer a Karush en persona años antes pero éste nunca le habló sobre su descubrimiento. Tanto Kuhn como Tucker estaban de acuerdo en que su resultado ya había sido obtenido previamente (en concreto, 11 años antes) por Karush y supieron reconocer su mérito. Los matemáticos de la época que se dedicaban a la Programación no lineal no entendían como Karush no dio un paso al frente reclamando prioridad o, al menos, reconocimiento. Karush, por su parte, se defendió dando la siguiente explicación:

Eso no responde a la pregunta de por qué no mencioné mi trabajo en los próximos años cuando la Programación no lineal se asentó y floreció. La razón de hacer esto pasó por mi cabeza alguna vez que otra, pero me sentía bastante tímido y reservado sobre mi precoz trabajo y no creo que tenga la enorme necesidad de ser reconocido por ello. En cualquier caso, la tesis de máster permaneció olvidada hasta hace pocos años cuando Hestenes me recomendó echarle un vistazo otra vez para ver si no obtuvo su correspondiente lugar en la historia... Así que revisé de nuevo la tesis y su trabajo con Tucker. Concluí que ustedes dos han explotado y desarrollado la materia mucho mas profundamente que yo, que no había justificación para que yo anunciase al mundo: '¡Mirad lo que yo hice primero!'

Esta respuesta que dio Karush es totalmente correcta y sincera. Él obtuvo un resultado comparable con el de Kuhn y Tucker, pero no exploró la materia mas a fondo; incluso su trabajo cayó del lado del Cálculo de variaciones, ni siquiera de la Programación no lineal. Bajo la dirección de Bliss, el departamento de matemáticas de Chicago se convirtió básicamente en un programa de investigación sobre problemas muy concretos y definidos del Cálculo de variaciones, de manera que nadie estaba interesado en explorar las posibles aplicaciones del resultado de Karush.

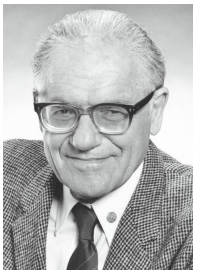


Figura 2.1: W. Karush, 1987

2.3. El Teorema de Fritz John: una contribución a la teoría de convexidad

La versión de F. John (1910-1994) del Teorema de Kuhn-Tucker apareció en su tratado *Extremum Problems with Inequalities as Subsidiary Conditions* [16], que fue publicado en un volumen de textos matemáticos con motivo de la celebración del 60º cumpleaños del matemático alemán Richard Courant en 1948.

John fue alumno de Courant en la Universidad de Gotinga donde realizó su tesis doctoral en 1933. Fue un gran matemático, llegando a trabajar para las Universidades de Cambridge, Kentucky y New York. Cuenta con 101 publicaciones matemáticas, que incluyen libros, artículos y monografías y ha sido galardonado con varios premios. Aunque es mayoritariamente conocido por su trabajo en ecuaciones diferenciales, también ha hecho importantes contribuciones en Geometría. En el momento de la publicación del volumen

en honor a Courant, John trabajaba en teoría de convexidad. De hecho, más de la mitad de sus publicaciones hasta 1948 eran en esta rama de las matemáticas y muchas de ellas son consideradas hoy día como textos clásicos dentro de esta teoría.

La intención del artículo de John quedó reflejada en la introducción del mismo donde escribió lo siguiente:

> *Este artículo trata sobre una extensión de la regla de los multiplicadores de Lagrange, donde las condiciones son desigualdades en vez de igualdades. Solamente se considerarán extremos de funciones diferenciables de un número finito de variables*

Al igual que Karush, John sólo está interesado en el caso finito-dimensional aunque más adelante en el trabajo, John apunta algo más lejos, diciendo lo siguiente

> *desde el punto de vista de las aplicaciones parece deseable extender el método usado aquí a casos donde las funciones involucradas no dependan de un número finito de variables independientes*

Esta extensión del problema de la que John habla pertenece claramente al Cálculo de variaciones. Es notable saber que, aparentemente, John no conocía la Escuela de Chicago ni sus contribuciones a este tipo de problemas. Así pues, ¿cuál era el verdadero interés de John con su artículo? La pregunta quedará resuelta mas adelante.

A continuación, se expone un breve esquema de lo que John hizo en su trabajo. En su artículo se diferencian dos partes: la primera parte trata sobre las condiciones necesarias y suficientes para la existencia de mínimo y la segunda parte trata sobre dos aplicaciones geométricas del resultado obtenido en la primera parte.

En la primera parte, John formuló el resultado relacionado con el Teorema de Kuhn-Tucker, que más tarde sería reconocido como una versión más débil de éste, como sigue:

> *Sea R un subconjunto de $\mathbb{R}^n$ y $F(\boldsymbol{x})$ una función real definida en R. Consideremos un subconjunto R' de R, el cual está descrito por un sistema de desigualdades con parámetro $\boldsymbol{y}$:*
>
> $$G(\boldsymbol{x}, \boldsymbol{y}) \geq 0$$
>
> *donde G es una función definida para todo $\boldsymbol{x}$ de R y todos los valores del parámetro $\boldsymbol{y}$. Asumamos que esos valores del parámetro $\boldsymbol{y}$ varían en un conjunto de puntos S en un espacio H. Estamos interesados en las condiciones que un punto $\boldsymbol{x^0}$ de R' tiene que verificar para que*
>
> $$M = F(\boldsymbol{x^0}) = \min_{\boldsymbol{x} \in R'} F(\boldsymbol{x})$$

Bajo ciertas condiciones de continuidad y diferenciabilidad, John fue capaz de demostrar el siguiente teorema:

Sea $\boldsymbol{x^0}$ un punto interior de R y perteneciente al conjunto R' de puntos de R que satisfacen las restricciones $G(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ para todo $\boldsymbol{y} \in S$. Sea

$$F(\boldsymbol{x^0}) = \min_{\boldsymbol{x} \in R'} F(\boldsymbol{x})$$

entonces existe un conjunto finito de puntos $y^1, ..., y^s \in S$ y números $\lambda_1, ..., \lambda_s$ no todos nulos tales que

$$\begin{cases} G(\boldsymbol{x^0}, \boldsymbol{y^r}) = 0 \quad \forall r = 1, ..., s \\ \lambda_0 \geq 0, \lambda_1 > 0, ..., \lambda_s > 0 \\ 0 \leq s \leq n \end{cases}$$

y la función $\phi(\boldsymbol{x}) = \lambda_0 F(\boldsymbol{x}) - \sum_{r=1}^{s} \lambda_r G(\boldsymbol{x}, \boldsymbol{y^r})$ tiene un punto crítico en $\boldsymbol{x}$

La manera en la que John atacó el problema fue la misma que la de Karush, pero donde Karush usaba el Lema de Farkas como su principal herramienta, John usó otros resultados similares procedentes de la teoría de la convexidad con la que se sentía más cómodo.

La formulación del Teorema de John parece algo distinta a la correspondiente de Karush, pero las condiciones que aparecen son esencialmente las mismas que las de Kuhn y Tucker. La diferencia reside en la aparición del parámetro $\boldsymbol{y}$ y el conjunto paramétrico S donde se mueve $\boldsymbol{y}$, y el hecho de que el multiplicador asociado a la función objetivo F puede ser cero (al igual que en la versión del Teorema de Karush). Esta última diferencia es causada por el hecho de que en el Teorema de John no se impone la *constraint qualification* (como dirían Kuhn y Tucker) o la condición de normalidad (como diría Karush).

En la segunda parte, que consiste en la aplicación del resultado a dos problemas geométricos, queda claro porqué John introduce el parámetro $\boldsymbol{y}$ y el conjunto paramétrico S. También se explica porque no consideró el problema incluyendo la *constraint qualification* o la condición de normalidad, que evitan que el multiplicador asociado a la función objetivo sea nulo.

Las aplicaciones geométricas de su resultado teórico trataban sobre la menor esfera conteniendo un conjunto dado y sobre la elipsoide de menor volumen conteniendo un conjunto dado. Mas concretamente, en la primera aplicación, John consideró el siguiente problema:

Sea S un subconjunto acotado de $\mathbb{R}^m$. Encontrar la esfera de menor radio positivo que encierra a S

John no estaba interesado en la existencia de dicha esfera. Si se asume que el conjunto acotado S contiene al menos dos puntos distintos, es bien claro que tal esfera existe.

Para poder usar su teorema, John caracterizó las esferas en $\mathbb{R}^m$ como puntos de $\mathbb{R}^{m+1}$, $\boldsymbol{x} = (x_1, ..., x_{m+1})$, donde $(x_1, ..., x_n)$ son las coordenadas de su centro y x_{m+1} es la raíz de su radio. Entonces, pudo reescribir el problema en cuestión como un problema de optimización sujeto a restricciones de desigualdad en la siguiente forma

Minimizar la función $F(\boldsymbol{x}) = x_{m+1}$ *sujeta a las restricciones*
$G(\boldsymbol{x}, \boldsymbol{y}) = x_{m+1} - \sum_{i=1}^{m}(x_i - y_i)^2 \geq 0$ *para todo* $\boldsymbol{y} \in S$

La restricción G garantiza que el mínimo de F se busca entre las esferas que contienen a S. John usó un procedimiento similar en la segunda aplicación sobre el elipsoide. En ambos casos, sabía que el mínimo existía, luego las condiciones necesarias de existencia se cumplían y usó esas condiciones para obtener importantes propiedades de la mínima esfera y mínimo elipsoide de las que dedujo mas adelante varias propiedades generales sobre conjuntos convexos y cerrados. En el fondo, su principal interés eran estos resultados sobre conjuntos convexos y cerrados y no el resultado teórico de la primera parte, que era una extensión del Teorema de Lagrange.

Según Kuhn, John debió haber revelado que obtuvo su teorema para después deducir que

> *la frontera de un conjunto compacto* S *de* $\mathbb{R}^n$ *permanece entre dos elipsoides homotópicas de radio menor o igual que* n *y que la elipsoide externa puede ser la elipsoide de menor volumen conteniendo a* S

Incluso pensando en el título e introducción, John da la impresión de estar interesado en problemas del Cálculo de variaciones, su artículo debe verse como una aportación a la teoría de la convexidad donde ha hecho grandes contribuciones. La conclusión es que lo que más le preocupaba a John eran las aplicaciones del resultado teórico de la primera parte y los resultados sobre convexidad que se podían derivar de ellas.

Por otro lado, el teorema que demuestra Karush en su trabajo era importante en sí mismo. La finalidad del artículo era obtener condiciones necesarias para la existencia de mínimo o máximo. En cambio, en el trabajo de John, el teorema es usado únicamente como herramienta para deducir resultados generales sobre conjuntos convexos. Estas aplicaciones explican la formulación con el parámetro $\boldsymbol{y}$ y el conjunto paramétrico S de su teorema. En cuanto a la carencia de la *constraint qualification* o condición de normalidad, se explica porque en ambas aplicaciones geométricas se verifica la *constraint qualification* y no es necesario exigirla como hipótesis en su resultado teórico de la primera parte.

En [29], Kuhn escribe sobre el trabajo de John que *está muy cerca de unirse al rango de textos clásicos y no publicados en nuestra materia* pero, al igual que Karush, John no concebió su trabajo dentro de la Programación no lineal y por eso nunca reclamó prioridad ni reconocimiento en este ámbito.

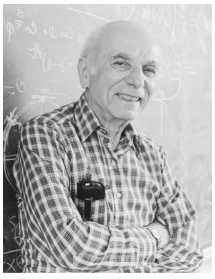


Figura 2.2: F. John, 1987

2.4. El Teorema de Kuhn y Tucker: una extensión de la Programación lineal

Albert W. Tucker nació en Canadá en 1905 y falleció en Princeton (New Jersey) en 1995, se graduó en matemáticas en 1928 por la Universidad de Toronto y un año después empezó su tesis doctoral en la Universidad de Princeton. En 1932, finalizó su tesis doctoral (sobre topología) y dos años más tarde se convirtió en profesor ayudante. En 1938 se hizo profesor asociado, y finalmente en 1946, llegó a ser profesor titular. Fue considerado una figura de importancia en investigación matemática que mantuvo en prestigio la Universidad de Princeton durante los años 30 y 40, siendo jefe del departamento de matemáticas entre 1953 y 1963 y se caracterizaba por ser buen profesor y tener una enorme influencia sobre todos sus alumnos.

Harold W. Kuhn nació en California en 1925 (20 años más joven que Tucker) y falleció en Nueva York en 2014 y se graduó en Ciencias en 1947 por el *Instituto de Tecnología de California*. Mas adelante, se trasladó a Princeton donde escribió una tesis titulada *Subgrup theorems for groups Presented by generators and Relations* en 1950 [37]. Tras 7 años, contratado en *Bryn Mawr College*, Kuhn regresó a Princeton como profesor asociado y estuvo en permanente contacto con los departamentos de matemáticas y economía. Fue íntimo amigo del matemático y Premio Nobel en Economía, John F. Nash.

El objetivo principal del artículo de Kuhn y Tucker [27] era encontrar condiciones necesarias y suficientes para la existencia de solución del siguiente problema de optimización al que denominaron *Problema del máximo*:

Encontrar un punto $\boldsymbol{x^0}$ *que maximice* $g(\boldsymbol{x})$
restringida por $\boldsymbol{F}(\boldsymbol{x}) \geq 0$ *para* $\boldsymbol{x} \geq \boldsymbol{0}$

Aquí $\boldsymbol{x^0} \in \mathbb{R}^n$ y $\boldsymbol{F} = (f_1, ..., f_m) : \mathbb{R}^n \to \mathbb{R}^m$ es una aplicación diferenciable, esto es, $\boldsymbol{F}(\boldsymbol{x})$ es un vector de dimensión m de componentes $f_1(\boldsymbol{x}), ..., f_m(\boldsymbol{x})$ que son funciones diferenciables definidas para todo $\boldsymbol{x} \geq \boldsymbol{0}$ y g es una función real, diferenciable y definida para todo $\boldsymbol{x} \geq \boldsymbol{0}$.

Kuhn y Tucker se enfrentaron a este problema tomando el denominado *Problema del punto de silla* como punto de partida. Este problema consistía en encontrar vectores no negativos $\boldsymbol{x^0} \in \mathbb{R}^n$ y $\boldsymbol{u^0} \in \mathbb{R}^m$ tales que

$$\phi(\boldsymbol{x}, \boldsymbol{u^0}) \leq \phi(\boldsymbol{x^0}, \boldsymbol{u^0}) \leq \phi(\boldsymbol{x^0}, \boldsymbol{u}) \qquad \forall \boldsymbol{x}, \boldsymbol{u} \geq 0$$

donde $\phi(\boldsymbol{x}, \boldsymbol{u})$ es una aplicación diferenciable de un n-vector de componentes $x_i \geq 0$ y un m-vector de componentes $u_j \geq 0$ y denotaron por $\boldsymbol{\phi^0_x}, \boldsymbol{\phi^0_u}$ los derivadas parciales de ϕ con respecto a $\boldsymbol{x}$ y $\boldsymbol{u}$ evaluadas en $\boldsymbol{x^0}$ y $\boldsymbol{u^0}$, respectivamente, es decir:

$$\boldsymbol{\phi^0_x} = \left(\frac{\partial \phi}{\partial x_1}(\boldsymbol{x^0}), ..., \frac{\partial \phi}{\partial x_n}(\boldsymbol{x^0})\right) \quad \text{y} \quad \boldsymbol{\phi^0_u} = \left(\frac{\partial \phi}{\partial u_1}(\boldsymbol{u^0}), ..., \frac{\partial \phi}{\partial u_m}(\boldsymbol{u^0})\right)$$

y usaron la notación $'$ para denotar el vector transpuesto.

El primer teorema que Kuhn y Tucker demuestran en su artículo trata sobre condiciones necesarias y suficientes para la existencia de solución del Problema del punto de silla. Ellos probaron que las condiciones

$$\boldsymbol{\phi}_{\boldsymbol{x}}^{\mathbf{0}} \leq \mathbf{0}, \qquad \langle (\boldsymbol{\phi}_{\boldsymbol{x}}^{\mathbf{0}})', \boldsymbol{x}^{\mathbf{0}} \rangle = 0, \qquad \boldsymbol{x}^{\mathbf{0}} \geq \mathbf{0} \tag{2.4}$$

$$\boldsymbol{\phi}_{\boldsymbol{u}}^{\mathbf{0}} \geq \mathbf{0}, \qquad \langle (\boldsymbol{\phi}_{\boldsymbol{u}}^{\mathbf{0}})', \boldsymbol{u}^{\mathbf{0}} \rangle = 0, \qquad \boldsymbol{u}^{\mathbf{0}} \geq \mathbf{0} \tag{2.5}$$

son necesarias para que $\boldsymbol{x}^{\mathbf{0}}, \boldsymbol{u}^{\mathbf{0}}$ proporcionen una solución del Problema del punto de silla. Para la segunda parte de la cuestión, probaron que las condiciones (2.4) y (2.5) junto con las condiciones

$$\phi(\boldsymbol{x}, \boldsymbol{u}^{\mathbf{0}}) \leq \phi(\boldsymbol{x}^{\mathbf{0}}, \boldsymbol{u}^{\mathbf{0}}) + \langle (\boldsymbol{\phi}_{\boldsymbol{x}}^{\mathbf{0}})', \boldsymbol{x} - \boldsymbol{x}^{\mathbf{0}} \rangle \tag{2.6}$$

$$\phi(\boldsymbol{x}^{\mathbf{0}}, \boldsymbol{u}) \geq \phi(\boldsymbol{x}^{\mathbf{0}}, \boldsymbol{u}^{\mathbf{0}}) + \langle (\boldsymbol{\phi}_{\boldsymbol{x}}^{\mathbf{0}})', \boldsymbol{u} - \boldsymbol{u}^{\mathbf{0}} \rangle \tag{2.7}$$

para todo $\boldsymbol{x}, \boldsymbol{u} \geq \mathbf{0}$ son suficientes.

Una vez llegados a este punto y equipados con estas condiciones, Kuhn y Tucker enunciaron su teorema de la siguiente manera:

Para que $\boldsymbol{x}^{\mathbf{0}}$ sea una solución del Problema del máximo es necesario que $\boldsymbol{x}^{\mathbf{0}}$ y $\boldsymbol{u}^{\mathbf{0}}$ verifiquen las condiciones (2.4) y (2.5) para $\phi(\boldsymbol{x}, \boldsymbol{u}) = g(\boldsymbol{x}) + \langle \boldsymbol{u}', \boldsymbol{F}(\boldsymbol{x}) \rangle$

Nótese que si se añade la condición $\boldsymbol{x}^{\mathbf{0}} \geq \mathbf{0}$ como restricción, entonces la primera y la última desigualdad de (2.4) implican que la función lagrangiana ϕ tiene un punto crítico en $(\boldsymbol{x}^{\mathbf{0}}, \boldsymbol{u})$. La segunda condición en (2.4) asegura que los multiplicadores asociados con las componentes no vinculadas de $\boldsymbol{x}^{\mathbf{0}}$ son iguales a cero. La primera condición en (2.5) asegura que $\boldsymbol{x}^{\mathbf{0}}$ es un punto factible, la segunda que los multiplicadores asociados a restricciones no vinculadas son iguales a cero y la última es la restricción del signo para los multiplicadores.

Estas condiciones se conocieron mas tarde como **condiciones de Kuhn y Tucker**, y constituyen uno de los resultados fundamentales en la teoría matemática de la Programación no lineal.

En realidad, la primera vez que Kuhn y Tucker anunciaron su teorema no fue en el *Simposio de Berkeley* sino unos meses antes en un seminario que tuvo lugar en *RAND Corporation* en mayo de 1950. Entre el público que asistió, se encontraba C. B. Tompkins, quien presentó un contraejemplo a su teorema y es que el resultado no podía descartar el caso en el que no se da la condición de normalidad, como lo hubiese llamado Karush. Kuhn y Tucker regresaron al trabajo y se percataron de que necesitaban ciertas condiciones de regularidad sobre las funciones restricciones. Esto les llevó a introducir la *constraint qualification*. La *constraint qualification* que ellos emplearon en su artículo fue la misma que la de Karush: a cada $\boldsymbol{x}^{\mathbf{0}}$ de la frontera del conjunto determinado por las restricciones y a cada vector diferencial $d\boldsymbol{x}$ para las cuales las derivadas direccionales de las correspondientes restricciones en la dirección de $d\boldsymbol{x}$ son no negativas, les corresponden un arco diferenciable $\boldsymbol{x} = a(\theta),\ 0 \leq \theta \leq 1$, contenido en el conjunto de las restricciones con $\boldsymbol{x}^{\mathbf{0}} = a(0)$ y un escalar postivo λ tal que $a'(\boldsymbol{x}^{\mathbf{0}}) = \boldsymbol{\lambda} d\boldsymbol{x}$. En la actualidad, las *constraint qualification* más comunes con las que uno tropieza al sumergirse en la Programación no lineal son la

regularidad o la superconsistencia (también denominada *condición de Slater*), las cuales son definidas en el tercer capítulo del trabajo. No obstante, éstas no son las únicas que hay.

Como Kuhn y Tucker señalaron en su escrito, puede parecer artificial introducir las condiciones (2.6) y (2.7) que tienen lugar para la suficiencia del Problema del punto de silla; pero estas condiciones son satisfechas si $\phi(\boldsymbol{x}, \boldsymbol{u^0})$ es una función cóncava de $\boldsymbol{x}$ y $\phi(\boldsymbol{x^0}, \boldsymbol{u})$ es una función convexa de $\boldsymbol{u}$. Para obtener una total equivalencia entre las soluciones del problema del máximo y el Problema del punto de silla entonces se requiere que las funciones involucradas $g, f_1, ..., f_m$ sean cóncavas y diferenciables. Con esta hipótesis extra, ellos probaron que

$\boldsymbol{x^0}$ es una solución del Problema del máximo sí, y sólo si, $\boldsymbol{x^0}$ y algún $\boldsymbol{u^0}$ proporcionan una solución del Problema del punto de silla para $\phi(\boldsymbol{x}, \boldsymbol{u}) = g(\boldsymbol{x}) + \langle \boldsymbol{u'}, \boldsymbol{F}(\boldsymbol{x}) \rangle$

Obviamente, la formulación del teorema por parte de Kuhn y Tucker es muy distinta a la de Karush y John, los cuales no consideraron el Problema del punto de silla. Pero, ¿por qué Kuhn y Tucker introducen el Problema del punto de silla como punto de partida para formular su teorema? y ¿por qué buscan una equivalencia entre el problema del máximo y el Problema del punto de silla?

La cooperación entre Kuhn y Tucker comenzó dos años antes en 1948, cuando examinaron la relación existente entre teoría de juegos y la Programación lineal que había sido desarrollada principalmente por George B. Dantzig para la *US Air Force*. Tucker, que era aun un estudiante, trabajó junto con otro compañero suyo David Gale, y ambos formularon el problema dual para un programa lineal general (véase la última sección del trabajo para mas información sobre Programación lineal) e hicieron notar la relación con la teoría de juegos. Ambos presentaron su descubrimiento en una conferencia que tuvo lugar en Chicago en junio de 1949. Pero entonces, Tucker siguió profundizando en el tema y resolvió un problema subyacente de minimización sobre la pérdida de calor. Según Kuhn, esto le permitió a Tucker familiarizarse con la regla de los multiplicadores de Lagrange, usada para problemas de optimización bajo restricciones de igualdad, y plantearse extender dicha regla para adaptarla a problemas de optimización bajo restricciones de desigualdad.

Tucker escribió a Gale y Kuhn para continuar la investigación en esa dirección. Gale rechazó la oferta pero Kuhn aceptó. Así que Kuhn y Tucker empezaron a trabajar con el propósito de extender el resultado dual de Programación lineal a Programación cuadrática, esto es, a problemas de optimización donde las funciones involucradas tienen forma cuadrática, mediante el clásico método de los multiplicadores de Lagrange.

Así pues, para resolver el problema de Programación lineal de maximizar la función $g(\boldsymbol{x}) = \sum c_i x_i$ con $c_i \in \mathbb{R}$ y donde $x_1, ..., x_n$ son n variables reales sujetas a $m + n$ las desigualdades lineales,

$$f_j(\boldsymbol{x}) = b_j - \sum a_{ji} x_i \geq 0, \qquad x_i \geq 0$$

Kuhn y Tucker formaron la correspondiente función lagrangiana

$$\phi(\boldsymbol{x}, \boldsymbol{u}) = g(\boldsymbol{x}) + \sum u_j f_j(\boldsymbol{x}), \qquad u_j \in \mathbb{R}$$

y notaron que $\boldsymbol{x^0} = (x_1^0, \ldots, x_n^0)$ maximizará g bajo las restricciones dadas si, y sólo si, existe un vector $\boldsymbol{u^0} = (u_1^0, \ldots, u_m^0) \in \mathbb{R}^m$ con componentes no negativas tal que $(\boldsymbol{x^0}, \boldsymbol{u^0})$ es un punto de silla para la función lagrangiana $\phi(\boldsymbol{x}, \boldsymbol{u})$.

En conclusión, un problema de Programación lineal tiene solución sí, y sólo si, la correspondiente función lagrangiana tiene un *punto de silla*. Además, este *punto de silla* constituye una solución no solo para el problema de Programación lineal sino también para el problema dual.

Considerando ahora que Kuhn y Tucker estaban buscando una forma de extender el Teorema dual para Programación lineal a casos más generales, parece perfectamente natural tomar el *punto de silla* para la función lagrangiana como punto de partida.

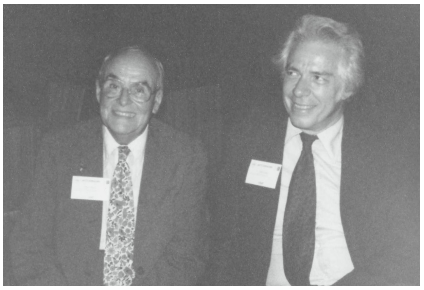


Figura 2.3: Kuhn (dcha.) y Tucker (izq.) en 1980

2.5. El aspecto de un descubrimiento múltiple

A parte de Karush, John y Kuhn y Tucker, también se le atribuye el resultado al matemático ruso M. Ostrogradsky (1801-1862) y al matemático húngaro J. Farkas (1847-1930). En [19], se discute la posibilidad de que el resultado de Kuhn y Tucker no es mas que un redescubrimiento independiente de un teorema obtenido por Ostrogradsky en un artículo leído para la Academia francesa en 1834 y publicado 4 años después. En [32], se da un repaso a toda la teoría de optimización desarrollada hasta 1980 y, de nuevo, se da por hecho que las condiciones de Kuhn–Tucker aparecen por primera vez en la literatura matemática en los trabajos de Ostrogradsky y Farkas.

Por el contrario, la comunidad matemática no atribuye el Teorema de Kuhn y Tucker a Ostrogradsky y Farkas. En cambio, sí que se consideran los trabajos de Karush y John como artículos pertenecientes a la Programación no lineal y en la mayoría de libros de texto y artículos relacionados con el tema, otorgan el resultado a Karush, Kuhn y Tucker. Todo parece indicar que nos encontramos ante un descubrimiento múltiple. El Teorema de Kuhn y Tucker es a menudo renombrado como el Teorema de Karush-Kuhn-Tucker. También hay una versión del teorema conocida como Teorema de Fritz John [30].

Lo que es particularmente confuso es que las apariciones de los resultados de Karush, John y Kuhn y Tucker, que mas tarde la comunidad matemática consideró como el mismo (o casi el mismo, en el caso de John) y que fueron desarrollados con una diferencia de 11 años, fueran recibidos de manera tan distinta.

Los respectivos trabajos de Ostrogradsky y Farkas no tuvieron influencia directa en el desarrollo de la Programación no lineal, de hecho ellos se centraban más en la teoría de desigualdades. Ambos dedujeron el resultado en un ámbito de la Mecánica analítica, cosa que no es de extrañar pues tanto John como Kuhn y Tucker mencionaron explícitamente que sus trabajos, de una forma u otra, estaban conectados con el método de los multiplicadores de Lagrange. John escribió, directamente, en la introducción que el propósito de su trabajo era extender este método a problemas con restricciones de desigualdad. Tucker asoció la Programación no lineal con las Leyes de Kirchoff en circuitos eléctricos y tuvo la idea de que quizás el método de los multiplicadores de Lagrange podía adaptarse al caso de restricciones de desigualdad. Es más, el propio Lagrange desarrollo su método en 1788 como un método para encontrar el punto de equilibrio en sistemas mecánicos y sentenció las bases sobre su teoría de equilibrio en lo que ahora es conocido como el Principio del trabajo virtual y que tomó como axioma. Para más información sobre la relación del método de Lagrange con la Mecánica y la obtención de los resultados de Ostrodgrasky y Farkas véase la sección 5 de [2].

Una de las figuras centrales en la literatura de los descubrimientos múltiples en ciencias es Robert K. Merton. La teoría de Merton se basa en diez puntos que le sirven para concluir si determinados resultados aparentemente debidos a un único matemático se deben considerar múltiples y según Merton, el Teorema de Khun-Tucker es un descubrimiento triple. De hecho, para Merton la gran mayoría de descubrimientos científicos son potencialmente múltiples.

La teoría de Merton no ha sobrevivido sin ser discutida. En contraposición se encuentra Don Patinkin, quien criticó el criterio de Merton sobre descubrimientos múltiples en Ciencia. Patinkin se basa en sólo dos puntos a distinguir para diferenciar descubrimientos múltiples y no múltiples. Patinkin afirma que muchos resultados atribuidos a varios científicos son en realidad debidos a uno sólo. Pero por sorpresa, su criterio también da como resultado que el Teorema de Kuhn-Tucker es un descubrimiento triple... parece ser que en este caso nadie duda: el teorema es un hallazgo triple por parte de Karush, Kuhn y Tucker y lo correcto y justo es referirse a él como el **Teorema de Karush-Kuhn-Tucker**, escribiendo el apellido Karush en primer lugar, pues este lo descubrió 11 años antes que Kuhn y Tucker.

2.6. La importancia del contexto histórico

En esta sección, se tratará sobre el contexto matemático y social en el que nació el Teorema de Karush-Kuhn-Tucker. Como ya ha quedado reflejado, los matemáticos de hoy en día conciben los resultados de Karush y Kuhn-Tucker como el mismo y el resultado de John como una variante de éste. En un contexto matemático, los tres resultados son

analizados ignorando las diferencias obvias y centrándose exclusivamente en las similitudes y la comunidad matemática ve los tres teoremas independientemente del contexto en el que fueron desarrollados. Un análisis que, por el contrario, se centre en las diferencias en las tres formulaciones del resultado y tenga en cuenta el contexto de las ramas de la matemática donde fueron desarrollados puede proporcionar una explicación sobre las diferentes influencias en el desarrollo matemático y en la recepción que tuvieron estos tres teoremas en la comunidad matemática.

Para comprender la fama y reconocimiento casi instantáneo que tuvo el trabajo de Kuhn y Tucker uno debe entender su origen en Matemática aplicada y la importancia de la organización posguerra de la ciencia en US, ambos consecuencias de la 2ª Guerra Mundial.

Introducción. Antes de la Segunda Guerra Mundial, la Matemática aplicada no tenía buena reputación entre los matemáticos en US. Desde principios del siglo XX, este país fue testigo de un crecimiento abultado en la comunidad de matemáticos. El tipo de investigación que perseguían los matemáticos era principalmente en lo que se conocía como Matemática pura. Tan solo un reducido número de matemáticos se dedicaban a la Matemática aplicada. En ambientes académicos había una jerarquía entre matemáticos y, en general, los que se dedicaban a la Matemática aplicada no estaban clasificados en posiciones muy altas, pero esta situación cambió a raíz de la Segunda Guerra Mundial. Durante este periodo una enorme cantidad de científicos formaron parte del esfuerzo que trajo consigo la guerra. Muchos de ellos fueron contratados directamente por las Fuerzas armadas y la mayoría se organizaron por medio de la *Oficina de Investigación Científica y Desarrollo* (OICD) que fue fundada en mayo de 1941, liderada por Vannevar Bush y financiada por el Congreso. Pero no fue hasta 1943, año en el cual el *Comité de Matemática Aplicada* (CMA) se creó como una suborganización dentro de OICD, cuando los matemáticos se involucraron en gran medida con la guerra. Los matemáticos organizados a través de CMA trabajaron bajo contrato y esto supuso el lazo entre las Fuerzas armadas y las matemáticas. Esta actividad durante la guerra sirvió para estimular la implicación de los matemáticos profesionales en la resolución de problemas de Matemática aplicada y en algunos casos estos problemas supusieron la aparición de nuevas disciplinas en matemáticas.

El problema de programación de las Fuerzas armadas. El trabajo y esfuerzo sobre lo que se convertiría en Programación lineal comenzó en la guerra. El principal responsable de esto fue George B. Dantzig que fue contratado en 1941 por las Fuerzas armadas para trabajar en los llamados *métodos de planificación de programas* (una herramienta de las Fuerzas armadas para llevar a cabo enormes planteamientos logísticos).

El apoyo científico de la posguerra. El fin a la guerra significó el fin de OICD. La organización de Bush fue una organización de emergencia y desde un principio se vió que la OICD iba a desaparecer al terminar la guerra y había una idea común de que los científicos regresarían a sus respectivos deberes en universidades. También había una fuerte creencia de que América debía ser fuerte científicamente para ser fuerte militarmente.

Mucha gente estaba concienciada sobre la intensa relación entre la ciencia y el ejército. Es más, por petición del presidente Roosevelt, Bush preparó un plan para la organización de la investigación posguerra en 1945. Roosevelt enfatizó la necesidad de una agencia independiente, establecida por ley y dedicada a la investigación militar. En su informe *The endless frontier*, Bush pidió un suministro de dinero por parte del gobierno para la investigación en las universidades e industrias. Bush quería que el gobierno apoyase la investigación básica sin necesidad de contentar al ejército. La idea de Bush era establecer una fundación (*National Science Foundation*) consciente de la importancia de la investigación en universidades e industrias pero llevó su tiempo crear tal fundación. Entre tanto, la Marina fundó la *Oficina de Investigación Naval* (OIN) con el propósito de continuar las prácticas de investigación establecidas por OICD.

Hacia la Programación lineal. Las diferentes secciones militares también contrataron científicos por su propia cuenta. Dantzig fue otra vez contratado por las Fuerzas aéreas donde trabajó desde 1946 hasta 1952 como asesor matemático para el cuartel general de la Fuerza Aérea de los Estados Unidos. La tarea para lo que fue contratado era desarrollar una especie de "máquina analógica" que recibiese cualquier tipo de ecuaciones, reglas o datos y usara estos para generar un plan o programa consistente para las Fuerzas aéreas. No obstante, esta primera idea de máquina analógica fue rechazada y el trabajo se volcó hacia lo que hoy llamamos Programación lineal. En la primavera de 1947, las Fuerzas aéreas establecieron el proyecto SCOOP (*Scientific Computation of Optimum Programs*) donde Dantzig, Wood y Geisler fueron las principales figuras. El objetivo de este proyecto era doble: la construcción un modelo matemático para el problema de Programación lineal y el desarrollo y construcción de ordenadores potentes capaces de realizar los cálculos necesarios para ello. El modelo acabó con lo que se refleja en el siguiente problema matemático: minimizar una función lineal sujeta a ecuaciones e inecuaciones lineales. Originalmente, Dantzig lo denominó *Programación en estructura lineal* y desarrolló el *método simplex*, un método iterativo para obtener la solución óptima en problemas de Programación lineal [22]. Para más información sobre el origen de la Programación lineal véase [22].

La participación de John Von Neumann. Dantzig recibió el consejo de ponerse en contacto con el economista T. C. Koopmans y el matemático J. V. Neumann. Koopmans se esforzó mucho para introducir la Programación lineal, especialmente a economistas, pero fue la intervención de Neumann la crucial para el desarrollo posterior.
Neumann estaba involucrado con casi todo lo relacionado con matemáticas que apareció durante la guerra, fue miembro de algunas Asesorías de Ciencia Militar extranjeras y tuvo numerosos trabajos de consulta militar. En octubre de 1947, conoció a Dantzig en Princeton, fue la primera vez que Neumann oía hablar sobre Programación lineal y fue capaz de intuir la relación con la teoría de juegos. En 1944, publicó su libro *Theory of Games and Economic Behavior* junto con el economista austriaco-americano Oskar Morgenstern. Ambos modelos (el modelo de Programación lineal y el modelo de teoría de juegos) pueden ser formulados a través de desigualdades lineales. Según Dantzig, en ese encuentro Neumann mostró que un juego bipersonal de suma cero se puede reducir

a un problema de Programación lineal y conjeturó la relación inversa[1]. El interés de Neumann se recogió en una nota que escribió en noviembre de 1947, [13]. En ella trabajó en un problema de máximos de Programación lineal sujeto a restricciones lineales de desigualdad:

$$\text{maximizar} \quad \langle \boldsymbol{a}, \boldsymbol{x} \rangle$$

$$\text{sujeto a} \quad \boldsymbol{x}A \leq \boldsymbol{\alpha} \quad \text{y} \quad \boldsymbol{x} \geq \boldsymbol{0}$$

donde $\boldsymbol{a}$ y $\boldsymbol{x}$ son vectores n-dimensionales, A es una matriz $n \times m$ y $\boldsymbol{\alpha}$ es un vector m-dimensional. Neumann estuvo apunto de demostrar (apunto porque usó una versión incorrecta del Lema de Farkas) que si existe un valor máximo finito en un punto $\boldsymbol{x_0}$, verificando las restricciones $\boldsymbol{x_0}A \leq \boldsymbol{\alpha}$ y $\boldsymbol{x_0} \geq \boldsymbol{0}$, entonces existe un vector m-dimensional $\boldsymbol{\xi}$ tal que $A\boldsymbol{\xi} \geq \boldsymbol{a}$ y $\langle \boldsymbol{a}, \boldsymbol{x} \rangle \geq \langle \boldsymbol{\xi}, \boldsymbol{\alpha} \rangle$. Además, este vector $\boldsymbol{\xi}$ minimiza la función lineal $\langle \boldsymbol{\xi}, \boldsymbol{\alpha} \rangle$. Así, el resultado en esta nota de Neumann se puede interpretar como el Teorema dual para Programación lineal 18. Neumann no estableció esta conclusión ni formuló el problema dual, lo que si hizo fue introducir las llamadas hoy en día *variables duales*, aunque él no las llamó así. De esta nota no puede decidirse si Neumann estaba realmente preocupado por la relación entre el problema primario y dual. En cualquier caso, esta nota es la primera señal del desarrollo de la Programación lineal en una teoría matemática. Neumann supuso una enorme influencia en acelerar todo este proceso.

La división matemática de OIN. El trabajo de Kuhn y Tucker, el cual es consecuencia directa de estas circunstancias, tuvo lugar bajo el contrato con la matemática de la división de OIN. A Mina Rees, una persona muy influyente en matemáticas, que trabajó como asistente técnico para Warren Weaver (el lider de CMA), se le pidió, justo después de la guerra, instalar un programa de fines matemáticos. Incluso sabiendo que ella dudaba del éxito de este programa, tomó el liderazgo porque consideraba extremadamente importante para el próspero desarrollo en US estar activamente involucrada en el programa de OIN. El programa que preparó para la OIN era uno que ella ya había analizado con la mayoría de lideres matemáticos de entre los departamentos del país. Realmente, estaba preocupada en si dicho programa reflejaría lo que los matemáticos consideran que ayudaría y beneficiaría realmente a las matemáticas. La pregunta era, por supuesto, si la Marina apoyaría la investigación básica y específica en Matemáticas puras sin ningún efecto a cambio hacia la Marina.
Rees quería que el programa reforzara la investigación matemática en USA y que no fragmentara el campo. En 1948, el departamento de matemáticas de la OIN ya había estado funcionado algo mas de un año y Rees anunció que *la filosofía que ha determinado los proyectos de investigación matemática patrocinados por OIN es respaldada.* Rees enfatizó que la investigación básica en matemáticas debía considerarse importante y estaba recibiendo financiación de OIN. Al final, cuando llegó el dinero, todo acabó con que un 80 % del gasto total fue a parar a investigación en Matemática aplicada, matemática estadística, análisis numérico y computación.

[1]Un juego bipersonal de suma cero es un juego para dos jugadores en el que cada uno tiene que escoger entre unas acciones dictadas a cada turno y la pérdida de un jugador supone el beneficio de su contrincante.

El apoyo de OIN. La expectativa de los métodos de planificación de programas que la Fuerza Aérea estaba desarrollando eran notables. En la primavera de 1948, Dantzig viajó a Princeton, en nombre de OIN, para encontrarse con Neumann y discutir las posibilidades para un proyecto universitario sobre Programación lineal y su relación con la teoría de juegos financiado por OIN. Durante esta visita, Dantzig conoció a Tucker y le introdujo brevemente en la Programación lineal. Tucker comentó la posible relación con las Leyes de Kirchoff y contactó con OIN unos días después preguntando si él encajaría en tal proyecto matemático.
Hasta el momento Tucker, que solo había dedicado su tiempo de investigación en Topología, aceptó convertirse en uno de los principales investigadores lo que cambió totalmente la dirección de sus investigaciones. Lo mismo ocurió con Kuhn, quien por el momento estaba acabando una tesis doctoral sobre teoría de grupos. En el verano de 1948, Kuhn llamó a Tucker preguntando por algún trabajo veraniego y Tucker le contrató, junto con David Gale, para trabajar en el proyecto de OIN. Ellos tres presentaron los resultados de sus trabajos en la primera conferencia sobre Programación lineal que tuvo lugar en Chicago en junio de 1949. El resultado mas destacado de entre todos fue el Teorema dual para Programación lineal. Tras esto, Kuhn y Tucker se comprometieron a fondo con el proyecto. El teorema dual despertó el interés de ambos y entonces intentaron extender el resultado a casos mas generales, lo que concluyó con el artículo de Programación no lineal y el Teorema de Kuhn-Tucker. Este trabajo fue también financiado por OIN.
Otro factor social también relacionado con el ejército fue el desarrollo de *Investigación Operativa* (IO) durante la guerra y el asentamiento de IO como disciplina científica en universidades tras la guerra. OIN jugó un papel importante en este proceso.
La Programación lineal fue incorporada inmediatamente a la "caja de herramientas" de la IO, lo que significó que la IO también estaba preparada para "acoger" la Programación no lineal tan pronto como fuese desarrollada. De este modo, puede parecer que OIN tuvo una influencia enorme en la creación de una comunidad matemática haciendo Programación lineal y en esta comunidad era casi inevitable que el artículo de Programación no lineal de Kuhn y Tucker adquiriera popularidad inmediata y diera lugar a una nuevo campo de investigación.
Durante las primeras dos décadas de su existencia, la Programación matemática se estableció por si sola como una disciplina a través de conferencias, monografías y libros de texto. En 1971, se creó la primera revista en el tema, *Mathematical Programming* y dos años mas tarde la *Mathematical Programming Society.*

2.7. Notas finales

El Teorema de Karush-Kuhn-Tucker enseña que un resultado matemático, en sí mismo, no siempre decide si producirá una larga investigación o no: el contexto social puede también jugar un papel importante. La importancia de un resultado y su potencial para estimular la investigación en la correspondiente área son determinados por el contexto matemático y, en ocasiones, social en el cual es desarrollado. El resultado de Kuhn-Tucker fue un resultado importante dentro de la disciplina matemática en la que Kuhn y Tucker estaban trabajando; una disciplina que también recibió un enorme apoyo financiero. Este no era el caso de las áreas donde los artículos de Karush y John surgieron.

El hecho de que Karush, John y Kuhn y Tucker recibieran todos reconocimiento por el teorema en la comunidad cientifica de la Programación no lineal se debe a la influencia de *third parties* –un concepto introducido por Susan Cozzens en [36]. En su libro, *Social Control and Multiples Discoveries in Science: The Opiate Receptor Case*, ella se centra en cómo un descubrimiento acaba considerándose al tiempo como múltiple y señala que, a menudo, es debido a un *after-the-fact process* donde el caso se resuelve por la influencia de *third parties*, es decir, miembros de la comunidad científica que no están involucrados de forma directa en el descubrimiento. A través de posteriores referencias y reconocimiento, la *third parties*, establece (casi de manera involutaria e inconsciente) el descubrimiento como múltiple.

El lector puede consultar más información sobre el origen y desarrollo de la Programación no lineal y, en especial, de su principal teorema en [2], [29] y [3].

Capítulo 3

El Teorema de Karush-Kuhn-Tucker y aplicaciones en Programación convexa

En este capítulo se exponen el enunciado y demostración del teorema fundamental de la Programación no lineal: el Teorema de Karush-Kuhn-Tucker, que generaliza al ya comentado Teorema de Lagrange del primer capítulo, y proporciona una condición necesaria de existencia de solución de un problema de optimización. Además, se tratará también un caso especial de programas: los programas convexos que tienen la peculiaridad de que la condición necesaria de existencia de solución dada por el Teorema de Karush-Kuhn-Tucker es también condición suficiente. Finalmente, se ilustrará la aplicabilidad del mencionado teorema con distintos ejemplos y se hará una introducción breve a la Programación lineal, geométrica y cuadrática, que caen en el dominio de la Programación convexa.

3.1. Programas con restricciones mixtas

Dados $n, p, q \in \mathbb{N}$ tres números naturales, un subconjunto no vacío Ω de $\mathbb{R}^n$ y $1+p+q$ funciones reales $f, g_1, ..., g_p, h_1, ..., h_q$ de clase $\mathcal{C}^1$ definidas en Ω, los problemas que se van a estudiar en este capítulo son los siguientes:

$$\left.\begin{array}{c} \text{minimizar } f(\boldsymbol{x}) \text{ sujeto a} \\ \begin{cases} g_1(\boldsymbol{x}) = 0, ..., g_p(\boldsymbol{x}) = 0 \\ h_1(\boldsymbol{x}) \leq 0, ..., h_q(\boldsymbol{x}) \leq 0 \end{cases} \\ \boldsymbol{x} \in \Omega \end{array}\right\} (PM^-) \qquad \left.\begin{array}{c} \text{maximizar } f(\boldsymbol{x}) \text{ sujeto a} \\ \begin{cases} g_1(\boldsymbol{x}) = 0, ..., g_p(\boldsymbol{x}) = 0 \\ h_1(\boldsymbol{x}) \leq 0, ..., h_q(\boldsymbol{x}) \leq 0 \end{cases} \\ \boldsymbol{x} \in \Omega \end{array}\right\} (PM^+)$$

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 $\{g_i(x) = 0 : i = 1, ..., p\}$ y desigualdades $\{h_j(x) \leq 0 : j = 1, ..., q\}$ se denominan *restricciones* de igualdad y desigualdad de (PM), respectivamente.

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 extremo local; si $p = 0$ y $q > 0$, entonces el problema (PM) incluye solo restricciones de desigualdad y el teorema original de Karush-Kuhn-Tucker (que solo consideraba restricciones de desigualdad) aporta una condición necesaria de existencia de extremo local; y si $p, q > 0$, entonces el problema (PM) incluye ambas restricciones y recibe el nombre de *programa de restricciones mixtas*. El caso extremo $p = q = 0$ da lugar a un problema de optimización libre ya estudiado en el Capítulo 1.

Definición 11 (Punto factible)**.** *Un* ***punto factible*** *para el problema (PM) es un punto de Ω que satisface todas las restricciones de (PM), es decir, un punto $\boldsymbol{x^0} \in \Omega$ tal que $g_i(\boldsymbol{x^0}) = 0$ y $h_j(\boldsymbol{x^0}) \leq 0$ para todo $i = 1, ..., p$ y $j = 1, ..., q$.*

Definición 12 (Región factible)**.** *La* ***región factible*** *del problema (PM) es el conjunto de todos los puntos factibles de (PM).*

Definición 13 (Problema consistente)**.** *El problema (PM) es un problema* ***consistente*** *si la región factible de (PM) es no vacía, esto es, si existe (al menos) un punto en Ω que satisface todas las restricciones de (PM).*

Definición 14 (Solución de (PM))**.** *Un mínimo (resp. máximo) local para el problema (PM^-) (resp. (PM^+)) es un mínimo (resp. máximo) local para el problema de minimizar (resp. maximizar) f en la región factible de (PM^-) (resp. (PM^+)) y una* ***solución*** *de (PM^-) (resp. (PM^+)) es un mínimo (resp. máximo) global de f en la región factible de (PM^-) (resp. (PM^+)).*

Definición 15 (Restricción activa e inactiva)**.** *Dado un punto factible $\boldsymbol{x^0} \in \Omega$ para el problema (PM), una* ***restricción activa*** *de (PM) en $\boldsymbol{x^0}$ es una restricción de (PM) que se anula en $\boldsymbol{x^0}$ y una* ***restricción inactiva*** *de (PM) en $\boldsymbol{x^0}$ es una restricción de (PM) que no se anula en $\boldsymbol{x^0}$. Evidentemente, las restricciones g_i para $i = 1, ..., p$ son todas activas en $\boldsymbol{x^0}$ y las restricciones h_j para $j = 1, ..., q$ son activas en $\boldsymbol{x^0}$ si $h_j(\boldsymbol{x^0}) = 0$ e inactivas si $h_j(\boldsymbol{x^0}) < 0$.*

Definición 16 (Punto regular)**.** *Un* ***punto regular*** *para el problema (PM) es un punto factible $\boldsymbol{x^0} \in \Omega$ de (PM) verificando que el conjunto de vectores siguiente es linealmente independiente*

$$\{\nabla g_i(\boldsymbol{x^0}), \nabla h_j(\boldsymbol{x^0}) \in \mathbb{R}^n : i \in \{1, ..., p\}, j \in J(\boldsymbol{x^0})\}$$

donde $J(\boldsymbol{x^0}) = \{k \in \{1, ..., q\} : h_k(\boldsymbol{x^0}) = 0\}$. Equivalentemente, un punto factible para (PM) es un punto regular para (PM) si el conjunto de vectores gradientes de las restricciones activas de (PM) en $\boldsymbol{x^0}$ es linealmente independiente.

Nota 6. *Nótese que bajo las hipótesis del Teorema de Lagrange, el punto $\boldsymbol{x^0}$ es regular.*

Para la definición anterior conviene distinguir dos casos particulares:

- *Caso sin restricciones ($p = q = 0$):* en un problema sin restricciones todos los puntos de Ω son regulares.
- *Caso con restricciones de desigualdad ($p = 0, q > 0$):* en un problema de restricciones de desigualdad, el punto $\boldsymbol{x^0} \in \Omega$ también es regular si $J(\boldsymbol{x^0}) = \emptyset$.

3.1.1. La demostración de McShane y ejemplos

El siguiente teorema proporciona una condición necesaria para la existencia de mínimo local para el problema (PM^-) y engloba los resultados de Lagrange y Karush, Kuhn-Tucker y John. Como consecuencia de este teorema se deducirá fácilmente la correspondiente condición de existencia de máximo local para el programa (PM^+).

Teorema 9 (Karush-Kuhn-Tucker generalizado para (PM^-))**.** *Sean $n, p, q \in \mathbb{N}$ tres números naturales, $\Omega \subset \mathbb{R}^n$ un subconjunto no vacío de $\mathbb{R}^n$, $f, g_1, ..., g_p, h_1, ..., h_q$ funciones reales de clase $\mathcal{C}^1$ en Ω y $\boldsymbol{x^0} \in \Omega^\circ$ un punto interior de Ω y factible para el problema (PM^-). Supongamos que $\boldsymbol{x^0}$ es solución del programa (PM^-), entonces existen números reales $\lambda_0, \lambda_1, ..., \lambda_p, \mu_1, ..., \mu_q \in \mathbb{R}$ no todos nulos tales que*

$$\lambda_0 D_k f(\boldsymbol{x^0}) + \sum_{i=1}^{p} \lambda_i D_k g_i(\boldsymbol{x^0}) + \sum_{j=1}^{q} \mu_j D_k h_j(\boldsymbol{x^0}) = 0 \qquad \forall k = 1, ..., n \tag{3.1}$$

Además,

i) $\lambda_0 \geq 0$ *y* $\mu_j \geq 0$ *para todo* $j = 1, ..., q$.

ii) Si $j \in \{1, ..., q\}$ *y* $h_j(\boldsymbol{x^0}) < 0$, *entonces* $\mu_j = 0$.

iii) Si $\boldsymbol{x^0}$ *es un punto regular es posible tomar* $\lambda_0 = 1$.

Demostración. Sin pérdida de generalidad asumimos que $\boldsymbol{x^0} = \boldsymbol{0}$, que $f(\boldsymbol{x^0}) = 0$ y que $h_1(\boldsymbol{x^0}) = 0, ..., h_z(\boldsymbol{x^0}) = 0, h_{z+1}(\boldsymbol{x^0}) < 0, ..., h_q(\boldsymbol{x^0}) < 0$ para algún $z \in \mathbb{N}$ con $1 < z < q$.

Por ser $\boldsymbol{x^0} = \boldsymbol{0}$ un punto interior de Ω, existe $\epsilon > 0$ tal que la bola abierta $B(\boldsymbol{0}, \epsilon)$ de centro el origen y radio ϵ está contenida en Ω y es inmediato que la bola cerrada $\overline{B}(\boldsymbol{0}, \epsilon_1)$ de centro el origen y radio $\epsilon_1 = \epsilon/2$ está contenida en Ω. Por el Teorema de conservación del signo, existe $\epsilon_2 > 0$ tal que las restricciones h_j con $j = z+1, ..., q$ son negativas en $\overline{B}(\boldsymbol{0}, \epsilon_2)$. Sea $\epsilon_0 = \text{mín}\{\epsilon_1, \epsilon_2\}$, entonces la bola cerrada $\overline{B}(\boldsymbol{0}, \epsilon_0)$ está contenida en Ω y las restricciones h_j para $j = z+1, ..., q$ son negativas en $\overline{B}(\boldsymbol{0}, \epsilon_0)$.

Lema 1. *Para cada $\epsilon > 0$ con $\epsilon \leq \epsilon_0$ existe algún $N \in \mathbb{N}$ tal que*

$$f(\boldsymbol{x}) + \|\boldsymbol{x}\|^2 + N \left(\sum_{i=1}^{p} g_i(\boldsymbol{x})^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x})^2 \right) > 0 \qquad \forall \boldsymbol{x} \in \mathbb{S}(\epsilon) \tag{3.2}$$

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

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

Existe $\tilde{\epsilon} \in (0, \epsilon_0]$ de forma que a cada $N \in \mathbb{N}$ le corresponde un punto $\boldsymbol{x_N} \in \mathbb{S}(\tilde{\epsilon})$ verificando $f(\boldsymbol{x_N}) + \|\boldsymbol{x_N}\|^2 + N \left(\sum_{i=1}^{p} g_i(\boldsymbol{x_N})^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x_N})^2 \right) \leq 0$

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 $\mathbb{S}(\tilde{\epsilon})$ de manera que

$$f(\boldsymbol{x_m}) + \|\boldsymbol{x_m}\|^2 + N_m \left(\sum_{i=1}^{p} g_i(\boldsymbol{x_m})^2 + \sum_{j=1}^{z} h_j^+(\boldsymbol{x_m})^2 \right) \leq 0 \qquad \forall m \in \mathbb{N} \tag{3.3}$$

Como $\{\boldsymbol{x_m}\}_{m\in\mathbb{N}}$ es una sucesión acotada de vectores de $\mathbb{R}^n$, el Teorema de Bolzano-Weiertrass garantiza la existencia de una sucesión parcial de $\{\boldsymbol{x_m}\}_{m\in\mathbb{N}}$ convergente a un punto $\boldsymbol{x^*} \in \mathbb{R}^n$. 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

$$\lim f(\boldsymbol{x_m}) = f(\boldsymbol{x^*}) \quad \text{y}$$

$$\|\boldsymbol{x^*}\| = \|\lim \boldsymbol{x_m}\| = \lim \|\boldsymbol{x_m}\| = \lim \tilde{\epsilon} = \tilde{\epsilon}$$

Dividiendo ahora ambos miembros de (3.3) por N_m obtenemos

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

Tomando límite cuando m tiende a ∞ en la expresión anterior y usando que $\lim N_m = \infty$, $\lim f(\boldsymbol{x_m}) = f(\boldsymbol{x^*})$ y $\lim \|\boldsymbol{x_m}\|^2 = \tilde{\epsilon}^2$, lo que implica $\lim \frac{f(\boldsymbol{x_m})}{N_m} = \lim \frac{\|\boldsymbol{x_m}\|^2}{N_m} = 0$, se sigue que

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

de donde se deduce que $\boldsymbol{x^*}$ satisface $g_i(\boldsymbol{x^*}) = 0$ para $i = 1, ..., p$ y $h_j^+(\boldsymbol{x^*}) = 0$ para $j = 1, ..., z$. Así que $\lim f(\boldsymbol{x_m}) = f(\boldsymbol{x^*}) \geq f(\boldsymbol{x^0} = \boldsymbol{0}) = 0$; pero por (3.3),

$$f(\boldsymbol{x_m}) \leq -\tilde{\epsilon}^2 < 0 \qquad \forall m \in \mathbb{N}$$

y de aquí se desprende que $\lim f(\boldsymbol{x_m}) = f(\boldsymbol{x^*}) < 0$. Contradicción. □

Lema 2. *Para cada $\epsilon > 0$ con $\epsilon \leq \epsilon_0$ existe un punto $\overline{\boldsymbol{x}} = (\overline{x}_1, ..., \overline{x}_n) \in \mathbb{R}^n$ y un vector unitario $(\lambda_0, \lambda_1, ..., \lambda_p, \mu_1, ..., \mu_z)$ con componentes $\lambda_0, \mu_1, ..., \mu_z$ no negativas tales que $\|\overline{\boldsymbol{x}}\| < \epsilon$ y*

$$\lambda_0 \left[D_k f(\overline{\boldsymbol{x}}) + 2\overline{x}_k \right] + \sum_{i=1}^{p} \lambda_i D_k g_i(\overline{\boldsymbol{x}}) + \sum_{j=1}^{z} \mu_j D_k h_j(\overline{\boldsymbol{x}}) = 0 \qquad \forall k = 1, ..., n \tag{3.4}$$

Demostración. Sea $\tilde{\epsilon} \in (0, \epsilon_0]$ fijo pero arbitrario y $N \in \mathbb{N}$ el número natural dado por el lema 1. Considérese la función $F : \overline{B}(\boldsymbol{0}, \tilde{\epsilon}) \subset \Omega \to \mathbb{R}$ definida por

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

Como F es continua y $\overline{B}(\mathbf{0},\tilde{\epsilon})$ compacto, el Teorema de Weiertrass asegura la existencia de un punto $\overline{\boldsymbol{x}} \in \overline{B}(\mathbf{0},\tilde{\epsilon})$ donde F alcanza su mínimo global, en particular $F(\overline{\boldsymbol{x}}) \leq F(\boldsymbol{x}^{\mathbf{0}} = \mathbf{0}) = 0$ y el lema 1 afirma que $\|\overline{\boldsymbol{x}}\| < \tilde{\epsilon}$ ($\overline{\boldsymbol{x}}$ es un punto interior de $B(\mathbf{0},\tilde{\epsilon})$). Así, en virtud del teorema 5, todas las derivadas parciales de primer orden de F deben anularse en $\overline{\boldsymbol{x}}$:

$$D_k f(\overline{\boldsymbol{x}}) + 2\overline{x}_k + 2N\left(\sum_{i=1}^{p} g_i(\overline{\boldsymbol{x}}) D_k g_i(\overline{\boldsymbol{x}}) + \sum_{j=1}^{z} h_j^+(\overline{\boldsymbol{x}}) D_k h_j(\overline{\boldsymbol{x}})\right) = 0 \qquad \forall k = 1, ..., n \quad (3.5)$$

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

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

para cualquier $k = 1, ..., n$, lo cual se demuestra en el capítulo 6 de [40] teniendo en cuenta que

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

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

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

es fácil comprobar que el vector $(\lambda_0, \lambda_1, ..., \lambda_p, \mu_1, ..., \mu_z)$ es unitario, λ_0 y μ_j $(j = 1, ..., z)$ son no negativos y dividiendo ambos miembros de (3.5) por τ se consigue (3.4). □

Finalmente, tomamos una sucesión decreciente de números reales y positivos $\{\delta_m\}_{m\in\mathbb{N}}$ teniendo a cero con $\delta_1 < \epsilon_0$. Para cada $m \in \mathbb{N}$, elegimos un punto $\overline{\boldsymbol{x}}_{\boldsymbol{m}} \in \mathbb{R}^n$ con $\|\overline{\boldsymbol{x}}_{\boldsymbol{m}}\| < \delta_m$ y un vector unitario $(\lambda_{0,m}, \lambda_{1,m}, ..., \lambda_{p,m}, \mu_{1,m}, ..., \mu_{z,m}, 0, ..., 0)$ con componentes $\lambda_{0,m}$ y $\mu_{j,m}$ $(j = 1, ..., z)$ no negativas tales que se cumple (3.4) (esto es posible por el lema 2). De nuevo, por el Teorema de Bolzano-Weiertrass, existe una sucesión parcial para la cual los vectores unitarios convergen a un límite $(\lambda_0, \lambda_1, ..., \lambda_p, \mu_1, ..., \mu_q)$. Como $\{\overline{\boldsymbol{x}}_{\boldsymbol{m}}\}_{m\in\mathbb{N}} \to \boldsymbol{x}^{\mathbf{0}} = \mathbf{0}$, la ecuación (3.4) se cumple para este vector límite y para $\boldsymbol{x}^{\mathbf{0}}$. Esto demuestra el teorema salvo el apartado *(iii)* que comentamos a continuación: si $\boldsymbol{x}^{\mathbf{0}}$ es un punto regular no puede ser $\lambda_0 = 0$ pues entonces (3.1) contradice la independencia lineal de los vectores $\nabla g_1(\boldsymbol{x}^{\mathbf{0}}), ..., \nabla g_p(\boldsymbol{x}^{\mathbf{0}}), \nabla h_1(\boldsymbol{x}^{\mathbf{0}}), ..., \nabla h_z(\boldsymbol{x}^{\mathbf{0}})$ y esto concluye la demostración del teorema. □

Figura 3.1: E. J. McShane

Esta demostración se le debe a Edward J. McShane (1904-1989) y fue publicada en un artículo [41] para *The American Mathematical Monthly* en el año 1973. Véase [14] para la biografía y contribuciones matemáticas de McShane. Curiosamente, en la demostración aportada por McShane no se necesita ninguna hipótesis sobre el número de restricciones (de igualdad y desigualdad) y el número de variables, a diferencia del Teorema de Lagrange que imponía, para poder aplicar el Teorema de la función implícita, que el número de restricciones (de igualdad, en este caso) fuese menor estricto que el número de variables de las que depende la función objetivo y dicha hipótesis queda eliminada con la demostración de McShane.

La demostración de McShane de este teorema es realmente asombrosa: únicamente aplica teoremas elementales del Análisis matemático como son el Teorema de Weiertrass, el Teorema de Bolzano-Weiertrass y la condición necesaria de existencia de extremo en un punto interior (teorema 5). La demostración es muy ingeniosa y aparentemente nada intuitiva, sin embargo, esto no es así: la idea que subyace de fondo en la prueba son los *métodos de funciones penalty*. McShane estaba muy familiarizado con las funciones penalty y los métodos consecuentes y ello le permitió establecer esta simple y bella demostración. Las funciones penalty se tratan de funciones F construidas a partir de la función objetivo f y restricciones de (PM) que cumplen ciertos requisitos (véase el capítulo 6 de [40]). Optimizando estas nuevas funciones, se puede obtener, bajo ciertas hipótesis, una sucesión de vectores de $\mathbb{R}^n$ tendiendo a la solución del problema (PM). La prueba de McShane no está muy extendida a pesar de su sencillez. La demostración más común, aunque bastante más compleja, puede verse en el capítulo 7 de [40].

No obstante, tres breves y simples demostraciones del mencionado teorema pueden verse también en [9], (donde se usa Álgebra lineal básica, el concepto de derivada de Fréchet, el Teorema de Weiertrass y la condición necesaria de existencia de extremo libre), en [7], (que es una prueba basada en la observación de que un cierto límite existe y es positivo y del Teorema de Weiertrass) o en [6], (donde se usa el Lema de Farkas y el Teorema de Weiertrass). Finalmente, destacamos también las demostraciones de B. H. Porciau [26] y a E. J. Beltrami [1].

Nota 7. *En la situación del Teorema de Karush-Kuhn-Tucker generalizado, las n ecuaciones (3.1) pueden reescribirse en forma vectorial como sigue*

$$\lambda_0 \nabla f(\boldsymbol{x^0}) + \sum_{i=1}^{p} \lambda_i \nabla g_i(\boldsymbol{x^0}) + \sum_{j=1}^{q} \mu_j \nabla h_j(\boldsymbol{x^0}) = 0 \tag{3.6}$$

Claramente, maximizar la función f equivale a minimizar la función $-f$. Así pues, basta aplicar el Teorema de Karush-Kuhn-Tucker generalizado a la función $-f$ para lograr una condición necesaria para la existencia de máximo local en el problema (PM^+). En concreto, el mencionado teorema aplicado a $-f$ concede la existencia de números reales $\lambda_0, \lambda_1, ..., \lambda_p, \mu_1, ..., \mu_q \in \mathbb{R}$ no todos nulos tales que

$$\lambda_0 D_k(-f(\boldsymbol{x^0})) + \sum_{i=1}^{p} \lambda_i D_k g_i(\boldsymbol{x^0}) + \sum_{j=1}^{q} \mu_j D_k h_j(\boldsymbol{x^0}) = 0 \qquad \forall k = 1, ..., n$$

pero estas ecuaciones se pueden escribir como

$$\lambda_0 D_k f(\boldsymbol{x^0}) + \sum_{i=1}^{p} (-\lambda_i) D_k g_i(\boldsymbol{x^0}) + \sum_{j=1}^{q} (-\mu_j) D_k h_j(\boldsymbol{x^0}) = 0 \qquad \forall k = 1, ..., n$$

y el Teorema de Karush-Kuhn-Tucker generalizado asegura entonces que $-\mu_j \geq 0$ para todo $j = 1, ..., q$ o, lo que es lo mismo, $\mu_j \leq 0$ para todo $j = 1, ..., q$. De aquí se desprende el siguiente resultado análogo al susodicho teorema para el máximo local del programa (PM^+):

Corolario 1 (Karush-Kuhn-Tucker generalizado para el problema (PM^+))**.** *En la situación del teorema 9, si $\boldsymbol{x^0}$ es solución del programa (PM^+), entonces existen números reales $\lambda_0, \lambda_1, ..., \lambda_p, \mu_1, ..., \mu_q \in \mathbb{R}$ no todos nulos tales que*

$$\lambda_0 D_k f(\boldsymbol{x^0}) + \sum_{i=1}^{p} \lambda_i D_k g_i(\boldsymbol{x^0}) + \sum_{j=1}^{q} \mu_j D_k h_j(\boldsymbol{x^0}) = 0 \qquad \forall k = 1, ..., n \tag{3.7}$$

Además,

i) $\lambda_0 \geq 0$ *y* $\mu_j \leq 0$ *para todo* $j = 1, ..., q$.

ii) Si $j \in \{1, ..., q\}$ *y* $h_j(\boldsymbol{x^0}) < 0$, *entonces* $\mu_j = 0$.

iii) Si $\boldsymbol{x^0}$ *es un punto regular es posible tomar* $\lambda_0 = 1$.

Al igual que pasaba con la condición de Lagrange, el ejemplo 7 sirve también para mostrar que el Teorema de Karush-Kuhn-Tucker generalizado se trata de una condición necesaria y no suficiente.

Nota 8. *Evidentemente, el Teorema de Lagrange es un caso particular del Teorema de Karush-Kuhn-Tucker generalizado. En el caso $q = 0$ (sólamente restricciones de igualdad son presentes), las condiciones i),ii) del teorema 9 carecen de sentido.*

Es común encontrar en la literatura matemática el uso del término **punto de Karush-Kuhn-Tucker** del problema (PM) para referirse a cualquier punto $\boldsymbol{x^0} \in \Omega$ verificando las llamadas **condiciones de Karush-Kuhn-Tucker**, es decir, que existen escalares $\lambda_0, \lambda_1, ..., \lambda_p, \mu_1, ..., \mu_q \in \mathbb{R}$, no todos nulos, satisfaciendo las siguientes condiciones:

Condición estacionaria: El punto en cuestión verifica las ecuaciones (3.1) o, igualmente, la ecuación 3.6

$$\lambda_0 D_k f(\boldsymbol{x^0}) + \sum_{i=1}^{p} \lambda_i D_k g_i(\boldsymbol{x^0}) + \sum_{j=1}^{q} \mu_j D_k h_j(\boldsymbol{x^0}) = 0 \qquad \forall k = 1, ..., n$$

Condición de factibilidad: El punto en cuestión es factible para el problema (PM)

$$g_i(\boldsymbol{x^0}) = 0 \quad \forall i = 1, ..., p$$

$$h_j(\boldsymbol{x^0}) \leq 0 \quad \forall j = 1, ..., q$$

Condición de holgura: El punto en cuestión verifica el apartado $ii)$ del teorema 9

$$\exists j \in \{1, ..., q\} : h_j(\boldsymbol{x^0}) < 0 \Rightarrow \mu_j = 0$$

Condición de signo: Si el punto en cuestión es un mínimo local para (PM^-), entonces $\mu_j \geq 0$ con $j = 1, ..., q$ y si es un máximo local para (PM^+), entonces $\mu_j \leq 0$ con $j = 1, ..., q$. Independientemente de si es mínimo o máximo local de (PM), siempre se tiene que $\lambda_0 \geq 0$ y no hay restricción de signo para los escalares $\lambda_1, ..., \lambda_p$ asociados a las restricciones de igualdad.

Los escalares $\lambda_1, ..., \lambda_p, \mu_1, ..., \mu_q$ se denominan *multiplicadores* y existe uno por cada restricción del programa. El multiplicador λ_i está asociado a la restricción de igualdad g_i para $i = 1, ..., p$ y el multiplicador μ_j está asociado a la restricción de desigualdad h_j para $j = 1, ..., q$. Como ya se dijo en el Capítulo 1, los valores $\lambda_1, ..., \lambda_p$ son llamados **multiplicadores de Lagrange** mientras que los valores $\mu_1, ..., \mu_q$ son llamados **multiplicadores de Karush-Kuhn-Tucker**.

Nótese que el apartado $i)$ del teorema 9 es la no negatividad de los escalares λ_0 y $\mu_1, ..., \mu_q$. Los multiplicadores de Lagrange carecen de restricción de signo (pueden ser positivos, cero y negativos); hecho que ya apareció con el Teorema de Lagrange en el Capítulo 1. El apartado $ii)$ del teorema 9 suele aparecer expresado de la siguiente manera: $\mu_j h_j(\boldsymbol{x^0}) = 0$ para todo $j = 1, ..., q$, que significa que si una restricción de desigualdad es inactiva en un punto, entonces el multiplicador de Karush-Kuhn-Tucker asociado a dicha restricción debe ser cero.

La función **lagrangiana** asociada al problema (PM) es la función $\mathcal{L} : \Omega \times \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ definida por la siguiente combinación lineal

$$\mathcal{L}(\boldsymbol{x}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \lambda_0 f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i g_i(\boldsymbol{x}) + \sum_{j=1}^{q} \mu_j h_j(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \Omega, \forall \lambda_0 \in \mathbb{R}, \forall \boldsymbol{\lambda} \in \mathbb{R}^p, \forall \boldsymbol{\mu} \in \mathbb{R}^q$$

donde $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_p)$ y $\boldsymbol{\mu} = (\mu_1, ..., \mu_q)$. O del mismo modo, en forma mas compacta:

$$\mathcal{L}(\boldsymbol{x}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \lambda_0 f(\boldsymbol{x}) + \langle \boldsymbol{\lambda}, g(\boldsymbol{x}) \rangle + \langle \boldsymbol{\mu}, h(\boldsymbol{x}) \rangle \quad \forall \boldsymbol{x} \in \Omega, \forall \lambda_0 \in \mathbb{R}$$

siendo $\boldsymbol{g} = (g_1, ..., g_p)$ y $\boldsymbol{h} = (h_1, ..., h_q)$.

Nota 9. *La condición estacionaria se puede expresar de manera mas reducida mediante la función lagrangiana recién introducida como* $\nabla_x \mathcal{L}(\boldsymbol{x^0}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0$.

En la práctica, para la búsqueda de puntos de Karush-Kuhn-Tucker (ya sean de mínimo o máximo) se han de seguir los siguientes pasos:

1. Construir la función lagrangiana asociada al problema (PM) y el sistema de ecuaciones compuesto por la condición estacionaria, la condición de factibilidad para las restricciones de igualdad y la condición de holgura:

$$\begin{cases} \lambda_0 D_k f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i D_k g_i(\boldsymbol{x}) + \sum_{j=1}^{q} \mu_j D_k h_j(\boldsymbol{x}) = 0 & \forall k = 1, ..., n \\ g_i(\boldsymbol{x}) = 0 & \forall i = 1, ..., p \\ \mu_j h_j(\boldsymbol{x}) = 0 & \forall j = 1, ..., q \end{cases}$$

2. Resolver el sistema (en general, no lineal) anterior de $n + p + q$ ecuaciones con $n+p+q+1$ incógnitas (que son las coordenadas del punto $(x_1, ..., x_n)$ y los multiplicadores $\lambda_0, \lambda_1, ..., \lambda_p, \mu_1, ..., \mu_q$). El procedimiento habitual de resolver este sistema es comenzar por la condición de holgura, que proporciona dos opciones

$$\mu_j h_j(\boldsymbol{x^0}) = 0 \Leftrightarrow \begin{cases} \mu_j = 0 \\ h_j(\boldsymbol{x^0}) = 0 \end{cases}$$

   para cada $j = 1, ..., q$. Así que, para q restricciones de desigualdad, se tienen 2^q casos posibles.

3. Una vez resuelto el sistema, hay que comprobar si los puntos obtenidos son puntos de Karush-Kuhn-Tucker: por una parte, hay que ver que son puntos factibles (en realidad, sólo es necesario comprobar si $h_j(\boldsymbol{x}) \leq 0$) y por otra parte, hay que ver que los multiplicadores de Karush-Kuhn-Tucker asociados tienen todos el mismo signo, o bien todos mayores o iguales que cero para mínimos locales o bien todos menores o iguales que cero para máximos locales.

Ejemplo 14.

$$\left.\begin{array}{c} \textit{optimizar } f(x, y, z) = x + y + z \textit{ sujeto a} \\ \begin{cases} h_1(x, y, z) = (y-1)^2 + z^2 \leq 1 \\ h_2(x, y, z) = x^2 + (y-1)^2 + z^2 \leq 3 \end{cases} \end{array}\right\}$$

Por el Teorema de Weiertrass, existe solución (máximo y mínimo) del programa considerado. La función lagrangiana $\mathcal{L} : \mathbb{R}^3 \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *asociada al programa es*

$$\mathcal{L}(x, y, z, \mu_1, \mu_2) = x + y + z + \mu_1[(y-1)^2 + z^2 - 1] + \mu_2[x^2 + (y-1)^2 + z^2 - 3]$$

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:

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

Condición de holgura:

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

Condición de signo:

$$\begin{cases} \mu_1, \mu_2 \geq 0 \longrightarrow \textit{mínimo local} \\ \mu_1, \mu_2 \leq 0 \longrightarrow \textit{máximo local} \end{cases}$$

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

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

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

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

$$P_3 = \left(\sqrt{2}, 1 + \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \qquad \boldsymbol{\mu} = \left(-\frac{1}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right)$$

$$P_4 = \left(\sqrt{2}, 1 - \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \qquad \boldsymbol{\mu} = \left(\frac{3}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}\right)$$

$$P_5 = \left(-\sqrt{2}, 1 + \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \qquad \boldsymbol{\mu} = \left(-\frac{3}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right)$$

$$P_6 = \left(-\sqrt{2}, 1 - \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \qquad \boldsymbol{\mu} = \left(\frac{1}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right)$$

Finalmente, se resume en una tabla todos los resultados obtenidos

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

3.2. Convexidad

La *convexidad* es una herramienta eficaz que aporta una condición suficiente a la hora de resolver ciertos problemas de optimización. Incluso, la convexidad permite sentar las bases para un importante proceso matemático conocido como *Programación convexa*. En esta sección se estudiará brevemente los conjuntos y funciones convexas (y cóncavas) y algunas de sus propiedades principales, únicamente con la finalidad de familiarizarnos con el concepto de convexidad que será crucial para el contenido de la parte final de este capítulo donde se verá la íntima relación que existe entre optimización y convexidad.

3.2.1. Conjuntos convexos

Comenzamos con la noción de *intervalo en varias variables*, que resulta esencial para comprender el concepto de conjunto convexo, seguida de la propia definición de conjunto convexo.

Definición 17 (Intervalo en varias variables)**.** *Sea $n \in \mathbb{N}$ un número natural y $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ dos puntos distintos de $\mathbb{R}^n$, entonces el* ***intervalo*** *$[\boldsymbol{x}, \boldsymbol{y}]$ entre $\boldsymbol{x}$ e $\boldsymbol{y}$ es el siguiente subconjunto de $\mathbb{R}^n$,*

$$[\boldsymbol{x}, \boldsymbol{y}] = \{\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \in \mathbb{R}^n : \lambda \in [0, 1]\}$$

Geométricamente, $[\boldsymbol{x}, \boldsymbol{y}]$ es el segmento de recta en $\mathbb{R}^n$ parametrizado por $\lambda \in [0, 1]$ que une el punto $\boldsymbol{x}$ con el punto $\boldsymbol{y}$. Obviamente, $[\boldsymbol{y}, \boldsymbol{x}]$ es el mismo conjunto que $[\boldsymbol{x}, \boldsymbol{y}]$ pero parametrizado en sentido opuesto.

Definición 18 (Conjunto convexo)**.** *Dado $n \in \mathbb{N}$ un número natural, un subconjunto Ω de $\mathbb{R}^n$ es* ***convexo*** *si para cada par de puntos $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ el intervalo $[\boldsymbol{x}, \boldsymbol{y}]$ está contenido en Ω. Simbólicamente,*

$$\Omega \text{ convexo} \Leftrightarrow \{\forall \boldsymbol{x}, \boldsymbol{y} \in \Omega \Rightarrow \lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \Omega \quad \forall \lambda \in [0,1]\}$$

Nótese que los papeles de $\boldsymbol{x}$ e $\boldsymbol{y}$ son intercambiables.

Nota 10 (Caso $n = 1$)**.** *En $\mathbb{R}$, la convexidad es equivalente a la conexión. Así, los subconjuntos convexos de $\mathbb{R}$ son justamente los intervalos de la recta real de cualquier tipo.*

Ejemplo 15 (Conjuntos convexos)**.** *En lo que sigue, n y m son números naturales cualesquiera.*

i) Si $\Omega, \Lambda \subset \mathbb{R}^n$ son subconjuntos convexos de $\mathbb{R}^n$ y $\alpha, \beta \in \mathbb{R}_0^+$ son números reales no negativos, entonces $\alpha\Omega + \beta\Lambda$ es un conjunto convexo de $\mathbb{R}^n$.

ii) Todo subespacio afin de $\mathbb{R}^n$ es convexo. Por ejemplo, las rectas, los planos o los hiperplanos afines de $\mathbb{R}^n$ son conjuntos convexos de $\mathbb{R}^n$.

iii) Dado $\boldsymbol{\xi} \in \mathbb{R}^n$ un punto de $\mathbb{R}^n$ y $\delta > 0$ un número real positivo, las bolas abierta y cerrada de centro $\boldsymbol{\xi}$ y radio δ son conjuntos convexos.

iv) Dada una familia de conjuntos convexos $\{\Omega_k \subset \mathbb{R}^n : k \in \mathbb{N}\}$, la intersección $\cap_{k\in\mathbb{N}}\Omega_k$ es un conjunto convexo de $\mathbb{R}^n$.

v) Sea $\phi : \mathbb{R}^n \to \mathbb{R}^m$ una aplicación lineal. Si $A \subset \mathbb{R}^n$ es convexo, entonces $\phi(A)$ es convexo en $\mathbb{R}^m$. Si $B \subset \mathbb{R}^m$ es convexo, entonces $\phi^{-1}(B)$ es convexo en $\mathbb{R}^n$.

vii) Dados $r, R \in \mathbb{R}^+$ números reales positivos con $r < R$ y $\boldsymbol{\xi} \in \mathbb{R}^n$ un punto de $\mathbb{R}^n$, el anillo $A(\boldsymbol{\xi}, r, R) = \{x \in \mathbb{R}^n : r < \|x - a\| < R\}$ de centro $\boldsymbol{\xi}$ y radios r y R no es un conjunto convexo.

Las demostraciones de estos ejemplos pueden verse en [40] y [10].

Dados $n, m \in \mathbb{N}$ dos números naturales, $\boldsymbol{x_1}, ..., \boldsymbol{x_m} \in \mathbb{R}^n$ m puntos de $\mathbb{R}^n$ y $\lambda_1, ..., \lambda_m \in \mathbb{R}$ m números reales no negativos cuya suma vale uno, se define la *combinación convexa* de $\boldsymbol{x_1}, ..., \boldsymbol{x_m}$ asociada a $\lambda_1, ..., \lambda_m$ como la combinación lineal dada por $\lambda_1\boldsymbol{x_1} + \cdots + \lambda_m\boldsymbol{x_m} \in \mathbb{R}^n$. El siguiente resultado establece que toda combinación convexa de un subconjunto convexo $\Omega \subset \mathbb{R}^n$ pertenece a Ω.

Teorema 10. *Sean $n, m \in \mathbb{N}$ dos números naturales, $\Omega \subset \mathbb{R}^n$ un subconjunto convexo de $\mathbb{R}^n$, $\boldsymbol{x_1}, ..., \boldsymbol{x_m} \in \Omega$ m puntos de Ω y $\lambda_1, ..., \lambda_m \in \mathbb{R}$ son números reales no negativos cuya suma vale uno, entonces la combinación convexa $\sum_{k=1}^m \lambda_k \boldsymbol{x_k}$ pertenece a Ω.*

La demostración se hace mediante inducción sobre m y puede verse en el Capítulo 2 de [40].

Dados $n \in \mathbb{N}$ un número natural y $\Omega \subset \mathbb{R}^n$ un subconjunto de $\mathbb{R}^n$, entonces existe el menor subconjunto convexo de $\mathbb{R}^n$ conteniendo a Ω, se denomina *envolvente conexa* de

Ω y se denota por $conv(\Omega)$. Matemáticamente, $conv(\Omega)$ es la intersección de todos los subconjuntos convexos de $\mathbb{R}^n$ conteniendo a Ω, entonces claramente $\Omega \subset conv(\Omega)$ y, por el ejemplo *iv*) anterior, $conv(\Omega)$ es un conjunto convexo de $\mathbb{R}^n$:

$$conv(\Omega) = \bigcap_{\substack{\Omega \subset \Lambda \\ \Lambda \text{ convexo}}} \Lambda$$

Un subconjunto $\Omega \subset \mathbb{R}^n$ es convexo si, y sólo si, $conv(\Omega) = \Omega$. Y por ejemplo, si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ son dos puntos de $\mathbb{R}^n$ y $\Omega = \{x, y\}$, entonces $conv(\Omega) = [\boldsymbol{x}, \boldsymbol{y}]$. Si $m \in \mathbb{N}$, $\boldsymbol{x_1}, ..., \boldsymbol{x_m} \in \mathbb{R}^n$ y $\Omega = \{\boldsymbol{x_1}, ..., \boldsymbol{x_m}\}$, entonces $conv(\Omega)$ es el poliedro convexo y cerrado de vértices $\boldsymbol{x_1}, ..., \boldsymbol{x_m}$. El siguiente resultado generaliza estos dos ejemplos.

Teorema 11. *Sea $n \in \mathbb{N}$ un número natural y $\Omega \subset \mathbb{R}^n$ un subconjunto de $\mathbb{R}^n$, entonces $conv(\Omega)$ coincide con el conjunto de todas las combinaciones convexas de puntos de Ω.*

Este teorema se demuestra probando que el conjunto de las combinaciones convexas $conv(\Omega)$ de Ω es un conjunto convexo conteniendo a Ω y que si C es un conjunto convexo conteniendo a Ω, entonces $conv(\Omega) \subset C$ para lo que se usa el teorema 10.

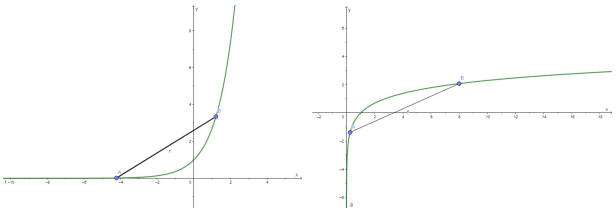
Corolario 2 (Caracterización de conjuntos convexos)**.** *Sea $n \in \mathbb{N}$ un número natural, entonces un subconjunto $\Omega \subseteq \mathbb{R}^n$ es convexo si, y sólo si, es cerrado para combinaciones convexas de puntos de Ω.*

Otras propiedades sobre conjuntos convexos pueden consultarse en [10], [18] y [12]. Destacamos, por ejemplo, que si $n \in \mathbb{N}$ es un número natural y $\Omega \subset \mathbb{R}^n$ es un subconjunto convexo de $\mathbb{R}^n$, entonces Ω° y $\overline{\Omega}$ también son convexos de $\mathbb{R}^n$. Además, si A es abierto (resp. acotado, compacto), entonces $conv(A)$ también es abierto (resp. acotado, compacto). Así como el siguiente resultado debido al matemático alemán C. Carathéodory, cuya demostración puede verse en [10].

Teorema 12 (Carathéodory)**.** *Sea $n \in \mathbb{N}$ un número natural, $A \subset \mathbb{R}^n$ un subconjunto de $\mathbb{R}^n$ y $a \in conv(A)$ un punto de la envolvente convexa de A, entonces a es combinación convexa de, a lo sumo, $n+1$ puntos de A.*

3.2.2. Funciones convexas

Empezamos con una interpretación geométrica de las funciones convexas y cóncavas reales de variable real. Si I es un intervalo no trivial de $\mathbb{R}$, entonces una función real $f : I \to \mathbb{R}$ definida en I será *convexa* (resp. *cóncava)* cuando la gráfica de la restricción de f a cualquier intervalo cerrado y acotado $[a, b]$ de I quede siempre contenida o por debajo (resp. por encima) del segmento de recta que une los puntos $(a, f(a))$ y $(b, f(b))$. Veamos esto con dos ejemplos concretos:



La función correspondiente a la gráfica de la izquierda es convexa y la función correspondiente a la gráfica de la derecha es cóncava. Dicho esto, conviene señalar (si aún el lector no lo ha hecho) que la función de la izquierda es la exponencial y la función de la derecha es el logaritmo.

Esta idea intuitiva sobre convexidad y cóncavidad se puede precisar y enunciar de forma rigurosa dando lugar a una definición para funciones reales de variable real que se extiende de manera natural a funciones de varias variables reales.

Definición 19 (Funciones convexas y cóncavas)**.** *Sea $n \in \mathbb{N}$ un número natural, $\Omega \subset \mathbb{R}^n$ un subconjunto convexo de $\mathbb{R}^n$ y $f : \Omega \to \mathbb{R}$ una función real definida en Ω, entonces*

- *f es **convexa** si $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ y $\lambda \in [0,1]$.*
- *f es **cóncava** si $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ y $\lambda \in [0,1]$.*
- *f es **estrictamente convexa** si $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) < \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$ para todo par de puntos $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ distintos y $\lambda \in (0,1)$.*
- *f es **estrictamente cóncava** si $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) > \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$ para todo par de puntos $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ distintos y $\lambda \in (0,1)$.*

Es fácil ver usando la definición anterior que si $\Omega \subset \mathbb{R}^n$ es un subconjunto convexo de $\mathbb{R}^n$ y $f : \Omega \to \mathbb{R}$ una función real definida en Ω, entonces

- si f es convexa, $epi(f) = \{(\boldsymbol{x}, \boldsymbol{y}) \in \Omega \times \mathbb{R} : \boldsymbol{y} \geq f(\boldsymbol{x})\} \subset \mathbb{R}^{n+1}$ (*epigrafo* de f) es un convexo de $\mathbb{R}^{n+1}$.
- si f es cóncava, $hipo(f) = \{(\boldsymbol{x}, \boldsymbol{y}) \in \Omega \times \mathbb{R} : \boldsymbol{y} \leq f(\boldsymbol{x})\} \subset \mathbb{R}^{n+1}$ (*hipografo* de f) es un convexo de $\mathbb{R}^{n+1}$.

Un primer resultado importante sobre funciones convexas y cóncavas es que toda función convexa o cóncava real de variable real definida en un intervalo abierto y no trivial de $\mathbb{R}$ es derivable por la izquierda y por la derecha y, en particular, continua (véase [17]).

Nota 11. *En la situación de la definición anterior, es claro que f es una función convexa si, y sólo si, $-f$ es una función cóncava.*

Por el teorema 10, se sabe que toda combinación convexa de una conjunto convexo permanece dentro del conjunto y esto da sentido al siguiente teorema, conocido como la *Desigualdad de Jensen* para funciones convexas.

Teorema 13. *Sean $n, m \in \mathbb{N}$ dos números naturales, $\Omega \subset \mathbb{R}^n$ un subconjunto convexo de $\mathbb{R}^n$, $\boldsymbol{x_1}, ..., \boldsymbol{x_m} \in \Omega$ m puntos de Ω y $f : \Omega \to \mathbb{R}$ una función real y convexa definida en Ω. Si $\lambda_1, ..., \lambda_m \in \mathbb{R}$ son números reales no negativos tales que $\lambda_1 + \cdots + \lambda_m = 1$, entonces*

$$f\left(\sum_{k=1}^{m} \lambda_k \boldsymbol{x_k}\right) \leq \sum_{k=1}^{n} \lambda_k f(\boldsymbol{x_k})$$

Si f es estrictamente convexa y $\lambda_1, ..., \lambda_m > 0$, entonces se da la igualdad arriba si, y sólo si, $\boldsymbol{x_1} = \cdots = \boldsymbol{x_m}$.

La demostración se hace por inducción sobre m y puede verse en el Capítulo 2 de [40].

Corolario 3. *Como consecuencia del teorema anterior y de la nota 11, se tiene que una función $f : \Omega \to \mathbb{R}$ cóncava en la misma situación del teorema 13 verifica la desigualdad contraria:*

$$f\left(\sum_{k=1}^{m} \lambda_k \boldsymbol{x_k}\right) \geq \sum_{k=1}^{n} \lambda_k f(\boldsymbol{x_k})$$

Dados $n \in \mathbb{N}$ un número natural y $x_1, ..., x_n \in \mathbb{R}^+$ números reales positivos, la *media aritmética* de $x_1, ..., x_n$ se define como el valor $\frac{1}{n}\sum_{k=1}^{n} x_k$ y la *media geométrica* de $x_1, ..., x_n$ se define como el valor $(\prod_{k=1}^{n} x_k)^{\frac{1}{n}}$. La versión más simple de la *desigualdad entre las medias aritmética y geométrica* establece que

$$\left(\prod_{k=1}^{n} x_k\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{k=1}^{n} x_k$$

y la igualdad se da si, y sólo si, $x_1 = \cdots = x_n$.

Nótese que los exponentes de los factores del miembro de la izquierda de (3.8) y los coeficientes de los sumandos del miembro de la derecha de (3.8) son iguales, positivos y suman uno. Estos números son llamados *pesos* asociados a las variables $x_1, ..., x_n$. La forma general de la desigualdad entre las medias aritmética y geométrica permite variar estos pesos siempre y cuando sean positivos y sumen uno y eso es lo que se demuestra en el siguiente ejemplo. En [40], puede verse la demostración de esta desigualdad mediante el método de los multiplicadores de Lagrange, pero resulta mucho mas breve y sencillo demostrarla haciendo uso de la convexidad de la función $-\log$ y del teorema anterior, como se muestra a continuación.

Corolario 4. *Sea $n \in \mathbb{N}$ un número natural, $x_1, ..., x_n \in \mathbb{R}^+$ números reales positivos y $\delta_1, ..., \delta_n \in \mathbb{R}^+$ números reales positivos cuya suma vale uno, entonces*

$$\prod_{k=1}^{n} (x_k)^{\delta_k} \leq \sum_{k=1}^{n} \delta_k x_k \tag{3.8}$$

y se da la igualdad si, y sólo si, $x_1 = \cdots = x_n$.

Demostración. La función $f : \mathbb{R}^+ \to \mathbb{R}$ dada por $f(x) = -\log(x)$ para todo $x \in \mathbb{R}$ es estrictamente convexa debido a que $f''(x) = 1/x^2 > 0$ para todo $x \in \mathbb{R}^+$. Consecuentemente, si $x_1, ..., x_n, \delta_1, ..., \delta_n \in \mathbb{R}^+$ números reales positivos tales que $\delta_1 + \cdots + \delta_n = 1$, el teorema 13 implica que

$$-\log\left(\sum_{k=1}^{n} \delta_k x_k\right) = f\left(\sum_{k=1}^{m} \delta_k x_k\right) \leq \sum_{k=1}^{n} \delta_k f(x_k) = -\sum_{k=1}^{n} \delta_k \log(x_k)$$

y, usando las propiedades de la función logaritmo, esta desigualdad equivale a

$$\log\left(\sum_{k=1}^{n} \delta_k x_k\right) \geq \sum_{k=1}^{n} \log(x_k^{\delta_k}) = \log\left(\prod_{k=1}^{n} x_k^{\delta_k}\right)$$

Como la función logaritmo es estrictamente creciente, se deduce que

$$\sum_{k=1}^{n} \delta_k x_k \geq \prod_{k=1}^{n} (x_k)^{\delta_k}$$

con igualdad si, y sólo si, $x_1 = \cdots = x_n$. □

Nota 12. *La desigualdad aritmético-geométrica fue probada por primera vez por el matemático francés Augustin-Louis Cauchy (1789-1857). Cauchy empleó una variante del principio de inducción para demostrar tal desigualdad. El lector puede consultar dicha demostración en el capítulo 2 de [15]. También puede demostrarse rápidamente a partir de razonamientos de convexidad con la función logaritmo (véase el corolario 4). Además, en [4] se demuestra que esta desigualadad es equivalente a la Desigualadad de Bernoulli. Cabe mencionar también que la Desigualdad de Young, a partir de la cual se demuestran la Desigualdad de Hölder y Minkowski, es consecuencia inmediata de la desigualdad aritmetico-geométrica (véase, por ejemplo, [40]).*

El siguiente teorema pone de manifiesto que las funciones convexas y cóncavas merecen un estudio mas a fondo a la hora de tratar con problemas de optimización.

Teorema 14. *Sea $n \in \mathbb{N}$ un número natural, $\Omega \subset \mathbb{R}^n$ un subconjunto convexo de $\mathbb{R}^n$ y $f : \Omega \to \mathbb{R}$ una función convexa definida en Ω, entonces todo mínimo local de f es también un mínimo global. Además, si f es estrictamente convexa en Ω, entonces el mínimo global es único.*

Demostración. Si $\boldsymbol{x^0}$ es un mínimo local de f, entonces existe $\delta \in \mathbb{R}^+$ tal que $f(\boldsymbol{x^0}) \leq f(\boldsymbol{x})$ para todo $\boldsymbol{x} \in B(\boldsymbol{x^0}, \delta)$. Dado $\boldsymbol{y} \in \Omega$ y $\lambda \in (0,1)$ con $\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x^0} \in \Omega$ y tal que $\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x^0} \in B(\boldsymbol{x^0}, \delta)$. Por hipótesis,

$$f(\boldsymbol{x^0}) \leq f(\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x^0}) \leq \lambda f(\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x^0})$$

Nótese que la última desigualdad es estricta si f es estrictamente convexa y $\boldsymbol{x^0} \neq \boldsymbol{y}$. Finalmente, de aquí se desprende que $f(\boldsymbol{x^0}) \leq f(\boldsymbol{y})$ con desigualdad estricta si f es estrictamente convexa y $\boldsymbol{x^0} \neq \boldsymbol{y}$. □

Corolario 5. *Como consecuencia del teorema anterior y de la nota 11, se tiene que todo máximo local de una función cóncava $f : \Omega \to \mathbb{R}$ es también un máximo global y si f es estrictamente cóncava en Ω, entonces el máximo global es único.*

Los siguientes resultados proporcionan una caracterización de las funciones convexas y cóncavas cuando se asume más regularidad y cuyas demostraciones pueden leerse en el Capítulo 2 de [40].

Proposición 1 (Caracterización para funciones de clase $\mathcal{C}^1$)**.** *Sea $n \in \mathbb{N}$ un número natural, $\Omega \subset \mathbb{R}^n$ un subconjunto abierto, convexo y no vacío de $\mathbb{R}^n$ y $f : \Omega \to \mathbb{R}$ una función real de clase $\mathcal{C}^1$ en Ω, entonces*

- *f es convexa si, y sólo 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 sólo 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 sólo 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 sólo 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 $\boldsymbol{x} \neq \boldsymbol{y}$.*

Corolario 6. *Sea $n \in \mathbb{N}$ un número natural, $\Omega \subset \mathbb{R}^n$ un subconjunto abierto, convexo y no vacío de $\mathbb{R}^n$ y f una función real y convexa (resp. cóncava) de clase $\mathcal{C}^1$ definida en Ω, entonces todo punto crítico de f en Ω es un mínimo (resp. máximo) global de f.*

Demostración. Supóngase que f es convexa y sean $\boldsymbol{x^0}, \boldsymbol{x} \in \Omega$ con $\boldsymbol{x^0}$ un punto crítico de f, entonces $\nabla f(\boldsymbol{x^0}) = \boldsymbol{0}$ y la proposición anterior implica que

$$f(\boldsymbol{x^0}) = f(\boldsymbol{x^0}) + \langle \nabla f(\boldsymbol{x^0}), \boldsymbol{x} - \boldsymbol{x^0} \rangle \leq f(\boldsymbol{x})$$

como se quería. La demostración para funciones cóncavas es análoga. □

Nota 13 (Caracterización para el caso $n = 1$ y derivable)**.** *Si $I \subset \mathbb{R}$ es un intervalo real no trival y $f : I \to \mathbb{R}$ es una función real derivable en I, entonces f es convexa (resp. cóncava) si, y sólo si, f' es creciente (resp. decreciente) en I. Véase [17] para la demostración. Además, si f' es estrictamente creciente (resp. estrictamente decreciente), entonces f es estrictamente convexa (resp. estrictamente cóncava). El recíproco no es cierto en general.*

Proposición 2 (Caracterización para funciones de clase $\mathcal{C}^2$)**.** *Sea $n \in \mathbb{N}$ un número natural, $\Omega \subset \mathbb{R}^n$ un subconjunto abierto, convexo y no vacío de $\mathbb{R}^n$ y $f : \Omega \to \mathbb{R}$ una función real de clase $\mathcal{C}^2$ en Ω, entonces*

- *f es convexa si, y sólo si, H_f es semidefinido positivo.*
- *f es cóncava si, y sólo si, H_f es semidefinido negativo.*

Nota 14. *Si H_f es definida positiva (resp. definida negativa), entonces f es estrictamente convexa (resp. estrictamente cóncava), pero es un error pensar que el recíproco es cierto.*

Nota 15 (Caracterización para el caso $n = 1$ y dos veces derivable)**.** *Si $I \subset \mathbb{R}$ es un intervalo real no trival y $f : I \to \mathbb{R}$ es una función real y dos veces derivable en I, entonces f es convexa (resp. cóncava) si, y sólo si, $f'' \geq 0$ (resp. $f'' \leq 0$) en I. Véase [17] para la demostración. Además, si $f'' > 0$ (resp. $f'' > 0$) en I, entonces f es estrictamente convexa (resp. estrictamente cóncava). El recíproco no es cierto en general.*

Proposición 3 (Álgebra con funciones convexas y cóncavas)**.** *En lo que sigue, n y m son números naturales cualesquiera.*

- *Si $f_1, ..., f_m$ son funciones reales y convexas (resp. cóncavas) definidas en un subconjunto convexo Ω de $\mathbb{R}^n$, entonces la función $F = f_1 + \cdots + f_m : \Omega \to \mathbb{R}$ es convexa (resp. cóncava). Además, si existe $k \in \{1, ..., m\}$ tal que f_k es estrictamente convexa (resp. estrictamente cóncava), entonces F es estrictamente convexa (resp. estrictamente cóncava). En general, el producto de funciones convexas (resp. cóncavas) no da como resultado una función convexa (resp. cóncava) y es fácil dar un contraejemplo de ello.*

- *Si f es una función real y convexa (resp. estrictamente convexa) definida en un subconjunto convexo Ω de $\mathbb{R}^n$, entonces si α es un número real positivo, la función $\alpha f : \Omega \to \mathbb{R}$ es convexa (resp. estrictamente convexa) y si β es un número real negativo, la función $\beta f : \Omega \to \mathbb{R}$ es cóncava (resp. estrictamente cóncava).*

- *Si f es una función real y convexa (resp. estrictamente convexa) definida en un subconjunto convexo Ω de $\mathbb{R}^n$ y g es otra función real, convexa y creciente (resp. estrictamente creciente) definida en un subconjunto $\Lambda \subset \mathbb{R}^n$ con $\Lambda \subset f(\Omega)$, entonces la composición $g \circ f$ es una función convexa (resp. estrictamente convexa). Análogamente, si f es una función real y cóncava (resp. estrictamente cóncava) definida en un subconjunto convexo Ω de $\mathbb{R}^n$ y g es otra función real, cóncava y decreciente (resp. estrictamente decreciente) definida en un subconjunto $\Lambda \subset \mathbb{R}^n$ con $\Lambda \subset f(\Omega)$, entonces la composición $g \circ f$ es una función cóncava (resp. estrictamente cóncava).*

Terminamos esta sección sobre convexidad aportando unos ejemplos de funciones convexas de varias variables.

Ejemplo 16. *En lo que sigue, n y m son números naturales cualesquiera.*

i) Toda función lineal definida en $\mathbb{R}^n$ es convexa y cóncava. Si $\boldsymbol{a} = (a_1, ..., a_n) \in \mathbb{R}^n$ y $b \in \mathbb{R}$, entonces es fácil comprobar que la función $f : \mathbb{R}^n \to \mathbb{R}$ definida como $f(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x} \rangle + b = a_1x_1 + \cdots + a_nx_n + b$ para todo $\boldsymbol{x} \in \mathbb{R}^n$ cumple que $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) = \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$ para cualesquiera $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ y $\lambda \in [0,1]$.

ii) La función exponencial $\exp : \mathbb{R} \to \mathbb{R}$ es convexa, la función logaritmo $\log : \mathbb{R}^+ \to \mathbb{R}$ es cóncava y la función potencia de exponente $\alpha \in \mathbb{R}$ es convexa cuando $\alpha \in (-\infty, 0] \cup [1, \infty]$ y cóncava cuando $\alpha \in [0,1]$ (nótese que en los casos $\alpha = 0$ y $\alpha = 1$, la función es convexa y cóncava).

iv) Sean $\boldsymbol{\alpha^{(1)}}, ..., \boldsymbol{\alpha^{(n)}} \in \mathbb{R}^n$ vectores fijos de $\mathbb{R}^n$ y $\beta_1, ..., \beta_m \in \mathbb{R}^+$ números reales positivos, entonces la función $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(\boldsymbol{x}) = \sum_{k=1}^m \beta_k e^{\langle \boldsymbol{\alpha^{(k)}}, \boldsymbol{x} \rangle}$ para todo $\boldsymbol{x} \in \mathbb{R}^n$ es convexa: como las funciones $g_k(\boldsymbol{x}) = \langle \boldsymbol{\alpha^{(k)}}, \boldsymbol{x} \rangle$ para $\boldsymbol{x} \in \mathbb{R}^n$ y $k = 1, ..., m$ son lineales y, por tanto, convexas y la función $h(t) = e^t$ para $t \in \mathbb{R}$ es convexa y estrictamente creciente, se tiene por la proposición 3 que las composiciones $h \circ g_k$ para $k = 1, ..., m$ son funciones convexas en $\mathbb{R}^n$ y como $\beta_1, ..., \beta_m$ son positivos, de nuevo la proposición 3 concluye que f es convexa.

v) Sean $\boldsymbol{\alpha^{(1)}}, ..., \boldsymbol{\alpha^{(n)}} \in \mathbb{R}^n$ vectores fijos de $\mathbb{R}^n$ y $\beta_1, ..., \beta_m \in \mathbb{R}^+$ números reales positivos, entonces la función $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(\boldsymbol{x}) = \log\left(\sum_{k=1}^m \beta_k e^{\langle \boldsymbol{\alpha^{(k)}}, \boldsymbol{x} \rangle}\right)$ para todo $\boldsymbol{x} \in \mathbb{R}^n$ es convexa. Computar el hessiano de f es complicado y resulta mucho mas fácil probar la convexidad de f haciendo uso de la Desigualdad de Hölder. Hemos de demostrar que $f(a\boldsymbol{x} + b\boldsymbol{y}) \leq af(\boldsymbol{x}) + bf(\boldsymbol{y})$ para cualesquiera $a, b \in \mathbb{R}^+$ con $a + b = 1$ y $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ o equivalentemente,

$$e^{f(a\boldsymbol{x}+b\boldsymbol{y})} \leq e^{af(\boldsymbol{x})+bf(\boldsymbol{y})} = (e^{f(\boldsymbol{x})})^a (e^{f(\boldsymbol{y})})^b \qquad \forall a, b \in \mathbb{R}^+ : a + b = 1, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

Sustituyendo la expresión de f, lo que hay que probar es

$$\sum_{k=1}^m \beta_k e^{\langle \boldsymbol{\alpha^{(k)}}, a\boldsymbol{x}+b\boldsymbol{y} \rangle} \leq \left(\sum_{k=1}^m \beta_k e^{\langle \boldsymbol{\alpha^{(k)}}, \boldsymbol{x} \rangle}\right)^a \left(\sum_{k=1}^m \beta_k e^{\langle \boldsymbol{\alpha^{(k)}}, \boldsymbol{y} \rangle}\right)^b$$

que es lo mismo que

$$\sum_{k=1}^{m}\left(\beta_k e^{\langle \boldsymbol{\alpha^{(k)}},x\rangle}\right)^a\left(\beta_k e^{\langle \boldsymbol{\alpha^{(k)}},y\rangle}\right)^b \leq \left(\sum_{k=1}^{m}\beta_k e^{\langle \boldsymbol{\alpha^{(k)}},x\rangle}\right)^a\left(\sum_{k=1}^{m}\beta_k e^{\langle \boldsymbol{\alpha^{(k)}},y\rangle}\right)^b$$

Sean $s_k = (\beta_k e^{\langle \boldsymbol{\alpha^{(k)}},\boldsymbol{x}\rangle})^a$ *y* $r_k = (\beta_k e^{\langle \boldsymbol{\alpha^{(k)}},\boldsymbol{y}\rangle})^b$ *para* $k = 1, ..., m$ *y* $a = 1/p$ *y* $b = 1/p^*$ *para algún* $p \geq 1$. *Con estas elecciones, la desigualdad de queremos probar es justamente la Desigualdad de Hölder,*

$$\sum_{k=1}^{m} s_k r_k \leq \left(\sum_{k=1}^{m} s_k^p\right)^{1/p}\left(\sum_{k=1}^{m} r_k^{p^*}\right)^{1/p^*}$$

vi) La función $f : \Omega = \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ *dada por*

$$f(x, y) = x^2 - 4xy + 5y^2 - \log(xy) \quad \forall (x, y) \in \Omega$$

es estrictamente convexa.
Basta escribir f *como la suma de dos funciones* $f = g + h$ *en* Ω *donde*

$$g(x, y) = x^2 - 4xy + 5y^2 \quad y \quad h(x, y) = -\log(xy) = -\log(x) - \log(y)$$

para todo $(x, y) \in \Omega$ *y razonar que* g *es estrictamente convexa por la proposición 2 puesto que* $H_g = \begin{pmatrix} 2 & -4 \\ -4 & 10 \end{pmatrix}$ *es definido positivo y también* h *es convexa ya que la función en una variable* $\varphi(t) = -\log(t)$ *en* $\mathbb{R}^+$ *es estrictamente convexa (porque* $\log$ *es estrictamente cóncava o si se prefiere porque* $\varphi''(t) = 1/t^2 > 0$ *para todo* $t \in \mathbb{R}^+$*), luego* $f = g + h$ *es estrictamente convexa por ser suma de dos funciones estrictamente convexas (proposición 3).*

Otras propiedades sobre funciones convexas pueden consultarse en [10], [18] y [12].

3.3. Programación convexa

Una de las áreas mas desarrolladas y estudiadas en la Programación no lineal es la *Programación convexa*, la cual se centra en la minimización de funciones convexas sujetas a restricciones de desigualdad también convexas. En esta sección se introducen los *programas convexos* que resultan de interés debido a que bajo hipótesis de convexidad, se obtiene una condición suficiente de existencia de solución (propia sólo de los programas convexos). Y, como cabe esperar, este resultado se puede trasladar a *programas cóncavos*, cuyo fin es maximizar una función cóncava sujeta a restricciones de desigualdad también cóncavas.

Dados $n, m \in \mathbb{N}$ dos números naturales, $\Omega \subset \mathbb{R}^n$ un subconjunto convexo y no vacío de $\mathbb{R}^n$ y funciones reales $f, h_1, ..., h_m$ convexas y de clase $\mathcal{C}^1$ definidas en Ω, el programa convexo que vamos a estudiar en esta sección es el siguiente problema de minimización restringida:

$$\left.\begin{array}{c}\text{minimizar } f(\boldsymbol{x})\\ \text{sujeto a } h_1(\boldsymbol{x}) \leq 0, ..., h_m(\boldsymbol{x}) \leq 0\\ \boldsymbol{x} \in \Omega\end{array}\right\} \quad (PC^-)$$

Dados $n, m \in \mathbb{N}$ dos números naturales, $\Omega \subset \mathbb{R}^n$ un subconjunto convexo y no vacío de $\mathbb{R}^n$ y funciones reales $f, g_1, ..., g_m$ cóncavas y de clase $\mathcal{C}^1$ definidas en Ω, el programa cóncavo que vamos a estudiar en esta sección es el siguiente problema de maximización restringida:

$$\left.\begin{array}{c}\text{maximizar } f(\boldsymbol{x})\\ \text{sujeto a } h_1(\boldsymbol{x}) \leq 0, ..., h_m(\boldsymbol{x}) \leq 0\\ \boldsymbol{x} \in \Omega\end{array}\right\} \quad (PC^+)$$

Escribiremos simplemente (PC) para referirnos indistintamente a (PC^+) o (PC^-).

Nota 16. *Es trivial que, dado un programa con restricciones de la forma $h_k \geq 0$ podemos transformarlo en uno de la forma (PC) sin mas que escribir $-h_k \leq 0$.*

De nuevo, las nociones de función objetivo, restricciones de desigualdad, punto factible, región factible, problema consistente y solución tienen el mismo sentido para (PC), aunque para este tipo de problemas se va a trabajar con un nuevo concepto: la *superconsistencia.*

Definición 20 (Programa superconsistente)**.** *El problema (PC) es* ***superconsistente*** *si existe un punto factible (**punto de Slater**) $\boldsymbol{x^0} \in \Omega$ para (PC) tal que $h_k(\boldsymbol{x^0}) < 0$ para todo $k = 1, ..., m$, es decir, en ninguna de las restricciones de desigualdad se da la igualdad al evaluar en $\boldsymbol{x^0}$.*

El siguiente ejemplo ayudará a aclarar estos conceptos.

Ejemplo 17.

$$\left.\begin{array}{c}\textit{minimizar } f(x,y) = x^4 + y^4\\ \textit{sujeto a } \begin{cases} h_1(x,y) = x^2 - 1 \leq 0\\ h_2(x,y) = y^2 - 1 \leq 0\\ h_3(x,y) = e^{x+y} - 1 \leq 0\end{cases}\\ (x,y) \in \mathbb{R}^2\end{array}\right\}$$

La región factible del programa es $\Lambda = \{(x,y) \in \mathbb{R}^2 : |x| \leq 1, |y| \leq 1, x + y \leq 0\} \neq \emptyset$, así que el programa es consistente. En realidad, es superconsistente puesto que cualquier punto interior a la región triangular determinada por las rectas $y = -x, y = -1$ y $x = -1$ es un punto de Slater. Como la función objetivo y las funciones restricción son convexas, el programa es convexo. En este caso, es fácil razonar sin hacer ningún cálculo que la única solución del programa es $(0,0)$.

Sin mas que añadir, se prueba ya la prometida condición suficiente en programas convexos.

Teorema 15 (Condición suficiente en Programación convexa)**.** *Sean $n, m \in \mathbb{N}$ dos números naturales, $\Omega \subseteq \mathbb{R}^n$ un subconjunto convexo y no vacío de $\mathbb{R}^n$, $f, h_1, ..., h_m : \Omega \to \mathbb{R}$ $m+1$ funciones reales, convexas y de clase $\mathcal{C}^1$ definidas en Ω, $\boldsymbol{x^0} \in \Omega^\circ$ un punto interior de Ω, factible y regular para el problema (PC^-). Supongamos que existen números reales $\lambda_0, \mu_1, ..., \mu_m \in \mathbb{R}$ no todos nulos satisfaciendo la condición estacionaria, de holgura y de signo (positivo), entonces $\boldsymbol{x^0}$ es solución de (PC^-), esto es, $\boldsymbol{x^0}$ es un mínimo global de f condicionado a $h_k(\boldsymbol{x}) \leq 0$ con $k = 1, ..., m$.*

Demostración. Como $\boldsymbol{x^0}$ es un punto regular por hipótesis, se puede tomar $\lambda_0 = 1$. Por hipótesis existen escalares $\mu_1, ..., \mu_m \in \mathbb{R}_0^+$ no todos nulos tales que $\nabla_x \mathcal{L}(\boldsymbol{x^0}) = 0$ donde $\mathcal{L} : \Omega \to \mathbb{R}$ es la función lagrangiana asociada a (PC^-), es decir,

$$\nabla f(\boldsymbol{x^0}) + \sum_{k=1}^{m} \mu_k \nabla g_k(\boldsymbol{x^0}) = 0$$

Como $\mu_k \geq 0$ para cada $k = 1, ..., m$ y $h_k(\boldsymbol{x}) \leq 0$ para cada $\boldsymbol{x} \in \Omega$ y $k = 1, ..., m$, se tiene que $\mu_k h_k(\boldsymbol{x}) \leq 0$ para cada $\boldsymbol{x} \in \Omega$ y $k = 1, ..., m$ y esto, junto con la proposición 1, permite escribir lo siguiente para cualquier punto factible $\boldsymbol{x} \in \Omega$,

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

□

Similarmente, se prueba el resultado análogo para el caso cóncavo.

Corolario 7 (Condición suficiente en Programación cóncava)**.** *Sean $n, m \in \mathbb{N}$ dos números naturales, $\Omega \subseteq \mathbb{R}^n$ un subconjunto convexo y no vacío de $\mathbb{R}^n$, $f, h_1, ..., h_m : \Omega \to \mathbb{R}$ $m+1$ funciones reales, cóncavas y de clase $\mathcal{C}^1$ definidas en Ω, $\boldsymbol{x^0} \in \Omega^\circ$ un punto interior de Ω, factible y regular para el problema (PC^+). Supongamos que existen números reales $\lambda_0, \mu_1, ..., \mu_m \in \mathbb{R}$ no todos nulos satisfaciendo la condición estacionaria, de holgura y de signo (negativo), entonces $\boldsymbol{x^0}$ es solución de (PC^+), esto es, $\boldsymbol{x^0}$ es un máximo global de f condicionado a $h_k(\boldsymbol{x}) \leq 0$ con $k = 1, ..., m$.*

Nota 17. *Evidentemente, los dos resultados anteriores son un caso particular del Teorema de Karush-Kuhn-Tucker generalizado, en el que $p = 0$ (sólamente restricciones de desigualdad son presentes).*

Obsérvese que se han usado todas las hipótesis en la demostración del teorema. Todas han jugado su papel en la prueba. No obstante, se puede reemplazar la hipótesis de regularidad del punto $\boldsymbol{x^0}$ por la superconsistencia, es decir, que $\boldsymbol{x^0}$ sea un punto de Slater.

De la proposición 1 se deduce también que si la función objetivo f de (PC^+) (resp. (PC^-)) es estrictamente convexa (resp. esrictamente cóncava), entones el programa (PC^+) (resp. (PC^-)) tiene una única solución.

Corolario 8. *En el programa (PC), las condiciones de Karush-Kuhn-Tucker son necesarias y suficientes para que un punto sea solución de (PC) (bajo las hipótesis de regularidad o superconsistencia).*

Antes de ver algunos ejemplos, es importante hacer hincapié en los siguientes hechos:

- Si $h_k : \Omega \to \mathbb{R}$ es una función convexa para algún $k \in \{1, ..., m\}$, entonces el conjunto $H_k = \{\boldsymbol{x} \in \Omega : h_k(\boldsymbol{x}) \leq 0\}$ es convexo. En particular, la región factible $\cap_{k=1}^m H_k$ de un programa convexo siempre será convexa por el ejemplo 15 apartado *iv)*.

- Para la demostración de la condición suficiente (teorema 15) basta con exigir que la función objetivo y las restricciones sean diferenciables y no de clase $\mathcal{C}^1$ en un subconjunto abierto de $\mathbb{R}^n$ como se imponía para la demostración de la condición necesaria (teorema 9). Las proposiciones 1 y 2 siguen siendo ciertas para funciones diferenciables y dos veces diferenciables, respectivamente.

Ejemplo 18. *Considere el programa*

$$\left.\begin{array}{l} \textit{minimizar } f(x,y) = (x-1)^2 + y^2 \\ \textit{sujeto a } \begin{cases} h_1(x,y) = x + y \leq 0 \\ h_2(x,y) = x^2 - 4 \leq 0 \end{cases} \end{array}\right\}$$

Este programa es un programa convexo y superconsistente al que se le puede aplicar el corolario 8. Así, el programa admite solución (x_0, y_0) si, y sólo si, existen multiplicadores $\mu_1, \mu_2 \in \mathbb{R}$ tales que

$$\begin{array}{l} \mu_1, \mu_2 \geq 0 \\ \mu_1(x_0 + y_0) = 0, \ \mu_2(x_0^2 - 4) = 0 \\ \begin{cases} 2x_0 - 2 + \mu_1 + 2\mu_2 x_0 = 0 \\ 2y_0 + \mu_1 = 0 \end{cases} \end{array}$$

No es difícil comprobar que el sistema anterior admite únicamente dos soluciones

$$(x_0, y_0) = (1, 0) \quad y \quad (\mu_1, \mu_2) = (0, 0)$$

$$(x_0, y_0) = \left(\frac{1}{2}, -\frac{1}{2}\right) \quad y \quad (\mu_1, \mu_2) = (1, 0)$$

pero la primera solución debe ser descartada como pues $(1, 0)$ no es un punto factible del programa. En cambio, $(1/2, -1/2)$ si es factible y el teorema 15 implica que es la única solución del programa.

Ejemplo 19. *Considere el programa*

$$\left.\begin{array}{l} \textit{minimizar } f(x,y) = x+y \textit{ sujeto a} \\ \begin{cases} h_1(x,y) = x^2+y^2 \leq 1 \\ h_2(x,y) = (x-1)^2+y^2 \leq 1 \end{cases} \end{array}\right\}$$

Se trata de un programa convexo y superconsistente al que se le puede aplicar el corolario 8. Así, el programa admite solución (x_0, y_0) si, y sólo si, existen multiplicadores $\mu_1, \mu_2 \in \mathbb{R}$ tales que

$$\begin{array}{l} \mu_1, \mu_2 \geq 0 \\ \mu_1(x_0+y_0) = 0, \ \ \mu_2(x_0^2-4) = 0 \\ \begin{cases} 1+2\mu_1 x_0 + 2\mu_2(x_0-1) = 0 \\ 1+2\mu_1 y_0 + 2\mu_2 y_0 = 0 \end{cases} \end{array}$$

cuya solución es $\mu_1 = 0, \mu_2 = \frac{\sqrt{2}}{2}$ y $(x_0, y_0) = (1-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$. Así, el valor mínimo que toma f condicionado a g_1 y g_2 es $f(x_0, y_0) = 1-\sqrt{2}$.

Ejemplo 20. *Considere el programa*

$$\left.\begin{array}{c} \textit{minimizar } f(x,y) = e^{-(x+y)} \\ \textit{sujeto a } \begin{cases} h_1(x,y) = e^x+e^y \leq 20 \\ h_2(x,y) = x \geq 0 \end{cases} \end{array}\right\}$$

Este programa es un programa convexo y superconsistente al que se le puede aplicar el corolario 8. Así, el programa admite solución (x_0, y_0) si, y sólo si, existen multiplicadores $\mu_1, \mu_2 \in \mathbb{R}$ tales que

$$\begin{array}{l} \mu_1, \mu_2 \geq 0 \\ \mu_1(e^{x_0}+e^{y_0}-20) = 0, \ \ \mu_2(-x_0) = 0 \\ \begin{cases} -e^{-(x_0+y_0)} + \mu_1 e^{x_0} - \mu_2 = 0 \\ -e^{-(x_0+y_0)} + \mu_1 e^{y_0} = 0 \end{cases} \end{array}$$

Es claro que $\mu_1 > 0$, pues si $\mu_1 = 0$ entonces la última ecuación diría que $e^{-(x+y)} = 0$. Así que hay dos posibilidades: $\mu_1, \mu_2 > 0$ y $\mu_1 > 0, \mu_2 = 0$.

Caso $\mu_1, \mu_2 > 0$:
En este caso, $x_0 = 0$ e $y_0 = \log 19$, pero $\mu_1 = 19^{-2} > 0$ y $\mu_2 = \mu_1 - 19^{-1} < 0$. No se cumple la condición de signo.

Caso $\mu_1 > 0, \mu_2 = 0$:
En este caso, $x_0 = y_0 = \log 10$ y $\mu_1 = 10^{-3} > 0$, luego el mínimo buscado es $f(x_0, y_0) = 10^{-2}$.

Ejemplo 21. *Considere el programa*

$$\left.\begin{array}{c} \textit{minimizar } f(x,y) = x^2+y^2-4x-4y \\ \textit{sujeto a } \begin{cases} h_1(x,y) = x^2-y \leq 0 \\ h_2(x,y) = x+y \leq 2 \end{cases} \end{array}\right\}$$

Este programa es un programa convexo y superconsistente al que se le puede aplicar el corolario 8. Así, el programa admite solución (x_0, y_0) *si, y sólo si, existen multiplicadores* $\mu_1, \mu_2 \in \mathbb{R}$ *tales que*

$$\begin{array}{l} \mu_1, \mu_2 \geq 0 \\ \mu_1(x_0^2 - y_0) = 0, \ \ \mu_2(x_0 + y_0 - 2) = 0 \\ \begin{cases} 2x_0 - 4 + 2\mu_1 x_0 + \mu_2 = 0 \\ 2y_0 - 4 - \mu_1 + \mu_2 = 0 \end{cases} \end{array}$$

La función objetivo es estrictamente convexa, luego hay unicidad de solución. Si, por ejemplo, $\mu_1 = 0$ *y* $\mu_2 > 0$ *entonces* $x_0 + y_0 = 2$ *y*

$$\begin{cases} 2x_0 - 4 + \mu_2 = 0 \\ 2y_0 - 4 + \mu_2 = 0 \end{cases}$$

Sumando ambas igualdades y usando que $x_0 + y_0 = 2$*, se llega a que* $\mu_2 = 2$ *y, por consiguiente, al punto* $(x_0, y_0) = (1, 1)$ *(factible) y es donde* f *alcanza su valor mínimo condicionado y vale* $f(1, 1) = -6$*. Como hay unicidad de solución, el problema queda completamente resuelto.*

La Programación convexa (resp. cóncava) se puede generalizar y considerar restricciones mixtas (de igualdad y desigualdad) siempre y cuando las restricciones de igualdad sean funciones lineales $g(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x} \rangle + b$, pues en ese caso el programa sigue siendo convexo (resp. cóncavo): la función objetivo y las restricciones de desigualdad son funciones convexas (resp. cóncavas), las restricciones de igualdad son funciones convexas (resp. cóncavas) por ser lineales y el dominio de definición de las funciones involucradas es un subconjunto convexo. Así que, el corolario 8 puede ser aplicado en estos casos igualmente.

Ejemplo 22. *Considere el programa*

$$\left.\begin{array}{c} \textit{optimizar } f(x, y, z) = y \textit{ sujeto a} \\ g(x, y, z) = x + y + z = 1 \\ h(x, y, z) = x^2 + z^2 \leq 9 \end{array}\right\}$$

Planteamos las condiciones de Karush-Kuhn-Tucker para la función lagrangiana $\mathcal{L} : \mathbb{R}^3 \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *dada por*

$$\mathcal{L}(x, y, z, \lambda, \mu) = y + \lambda(x + y + z - 1) + \mu(x^2 + z^2 - 9)$$

Condición estacionaria: $\begin{cases} \lambda + 2\mu x = 0 \\ 1 + \lambda = 0 \\ \lambda + 2\mu z = 0 \end{cases}$

Condición de factibilidad: $\begin{cases} x + y + z = 1 \\ x^2 + z^2 \leq 9 \end{cases}$

Condición de holgura: $\mu(x^2 + z^2 - 9) = 0$

Condición de signo: $\begin{cases} \mu \geq 0 \Longrightarrow & \textit{mínimo local del programa} \\ \mu \leq 0 \Longrightarrow & \textit{máximo local del programa} \end{cases}$

De la segunda ecuación de la condición estacionaria se tiene directamente que $\lambda = -1$ *y de la condición de holgura se tienen dos casos: o bien* $\mu = 0$ *o bien* $x^2 + z^2 = 9$*, pero la primera opción no es válida pues* $\mu \neq 0$ *ya que en otro caso, la primera ecuación de la condición estacionaria establece que* $\lambda = 0$ *lo cual es absurdo (*$\lambda = -1$*). Así pues,* $x^2 + z^2 = 9$ *y sustituyendo el valor* $\lambda = -1$ *en la condición estacionaria se obtiene el siguiente sistema*

$$\begin{cases} -1 + 2\mu x_0 = 0 & (1) \\ -1 + 2\mu z_0 = 0 & (2) \\ x_0 + y_0 + z_0 = 1 & (3) \\ x_0^2 + z_0^2 = 9 & (4) \end{cases}$$

Por una lado, de (1) y (2) y del hecho de que $\mu \neq 0$ *(pues si* $\mu = 0$*, entonces (1) y (2) conducen a que* $-1 = 0$*) se desprende que* $x_0 = z_0$ *y sustituyendo en (4),*

$$x_0 = z_0 = \pm\frac{3}{\sqrt{2}}$$

Por otro lado, el valor de y *se obtiene de (3),*

$$y_0 = 1 - x_0 - z_0 = 1 \mp \frac{6}{\sqrt{2}} = 1 \mp 3\sqrt{2}$$

Finalmente, el valor de μ *se obtiene de (1) o de (2),* $\mu = \pm\frac{1}{3\sqrt{2}}$*. Se han conseguido dos puntos:*

$$P = \left(\frac{3}{\sqrt{2}}, 1 - 3\sqrt{2}, \frac{3}{\sqrt{2}}\right) \ \textit{para} \ \left(\lambda = -1, \mu = \frac{1}{3\sqrt{2}}\right)$$

$$Q = \left(-\frac{3}{\sqrt{2}}, 1 + 3\sqrt{2}, -\frac{3}{\sqrt{2}}\right) \ \textit{para} \ \left(\lambda = -1, \mu = -\frac{1}{3\sqrt{2}}\right)$$

El punto P *tiene asociado un multiplicador de Karush-Kuhn-Tucker positivo y el punto* Q *tiene asociado un multiplicador de Karush-Kuhn-Tucker negativo. Como el programa es convexo,* P *es el mínimo global y* Q *es el máximo global de* f *condicionado a* g *y* h*.*

Ejemplo 23. *Considere el programa*

$$\left.\begin{array}{c} \textit{minimizar } f(x,y) = 2x^2 + 2xy + y^2 - 10x - 10y \\ \textit{sujeto a } \begin{cases} x^2 + y^2 \leq 5 \\ 3x + y \leq 6 \end{cases} \end{array}\right\}$$

Por el Teorema de Weiertrass, existe solución del programa anterior. Además, el programa es convexo pues tanto las restricciones como la función objetivo son convexas. De hecho, la función objetivo es estrictamente convexa ya que podemos escribir

$$f(x,y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - 10x - 10y$$

y observar que la matriz $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ *es definida positiva. Nótese que el programa es superconsistente (el origen es un punto de Slater para el programa). Las condiciones de Karush-Kuhn-Tucker son*

$$\begin{array}{l} \mu_1, \mu_2 \geq 0 \\ \mu_1(x_0^2 + y_0^2 - 5) = 0, \ \mu_2(3x_0 + y_0 - 6) = 0 \\ \begin{cases} 4x_0 + 2y_0 - 10 + 2\mu_1 x_0 + 3\mu_2 = 0 \\ 2x_0 + 2y_0 - 10 + 2\mu_1 y_0 + \mu_2 = 0 \end{cases} \end{array}$$

Supongamos que la primera restricción es activa, esto es, $\mu_1 > 0$ *y que la segunda restricción es inactiva, esto es,* $\mu_2 = 0$*, entonces las condiciones de KKT quedan*

$$\begin{cases} 4x_0 + 2y_0 - 10 + 2\mu_1 x_0 = 0 \\ 2x_0 + 2y_0 - 10 + 2\mu_1 y_0 = 0 \\ x_0^2 + y_0^2 = 5 \end{cases}$$

que tiene por solución $(x_0, y_0) = (1, 2)$ *y* $\mu_1 = 1$*. Como hay unicidad de solución, el problema queda completamente resuelto.*

En este trabajo se han presentado la potencia de las hipótesis de superconsistencia o regularidad, que son dos condiciones son dos ejemplos de *constraint qualifications* (hipótesis de cualificación de restricciones). No obstante, en la literatura matemática, también aparecen teorías correspondientes a otras *constraint qualifications*. Por ejemplo, el problema (PM) también puede ser estudiado satisfactoriamente mediante las **condiciones de Mangasarian-Fromowitz**. Estas *constraint qualifications* en un punto factible $\boldsymbol{x} \in \Omega$ pueden enunciarse como sigue

a) El conjunto $\{\nabla g_i(\boldsymbol{x}) : i = 1, ..., p\}$ es linealmente independiente.

b) Existe un vector $\boldsymbol{y^*} \in \mathbb{R}^n$ (llamado *vector de Mangasarian-Fromowitz*) tal que

$$\langle \nabla g_i(\boldsymbol{x}), \boldsymbol{y^*} \rangle = 0 \quad \forall i = 1, ..., p$$

$$\langle \nabla h_j(\boldsymbol{x}), \boldsymbol{y^*} \rangle < 0 \quad \forall j \in \{1, ..., q\} \text{ con } h_j(\boldsymbol{x}) = 0$$

Para más información sobre otros tipos de *constraint qualifications* véase [11] y [33].

Para concluir este capítulo y, con ello, este trabajo veamos diversas aplicaciones del Teorema de Karush-Kuhn-Tucker a la Programación convexa que derivarán en el estudio de tres famosos tipos de programas: la *Programación lineal*, *geométrica* y *cuadrática*.

3.4. Introducción a la Programación convexa dual y ejemplos

En lo que sigue, vamos a centrar la atención en programas convexos y, por comodidad, estos serán denotados simplemente por (PC) en vez de (PC^-). A cada programa convexo (PC) (*primal*) se le puede asociar un nuevo programa (PCD) (*dual*). El programa (PCD) es un programa de maximización no restringido que, a menudo, es mas fácil de resolver que el programa (PC) y cuyas soluciones pueden usarse para generar soluciones de (PC).

Dados $n, m \in \mathbb{N}$ números naturales y el siguiente programa convexo

$$\left.\begin{array}{c} \text{minimizar } f(\boldsymbol{x}) \text{ sujeto a} \\ h_1(\boldsymbol{x}) \leq 0, ..., h_m(\boldsymbol{x}) \leq 0 \\ \boldsymbol{x} = (x_1, ..., x_n) \in \Omega \end{array}\right\} \quad (PC)$$

donde las funciones $f, h_1, ..., h_m$ son funciones convexas y de clase $\mathcal{C}^1$ definidas en un subconjunto convexo Ω de $\mathbb{R}^n$, se define el **programa dual** (PCD) de (PC) como

$$\left.\begin{array}{c} \text{maximizar } g(\boldsymbol{\lambda}) = \inf\limits_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \\ \text{donde } \boldsymbol{\lambda} \geq \boldsymbol{0} \text{ y } \mathcal{L} \text{ es el lagrangiano} \\ \text{asociado al programa } (PC) \end{array}\right\} \quad (PCD)$$

Aquí, la notación $\boldsymbol{\lambda} \geq \boldsymbol{0}$ significa que todas las componentes del vector $\boldsymbol{\lambda} \in \mathbb{R}^m$ son mayores o iguales que cero.

Definición 21 (Supremo e ínfimo)**.** *Sea $n \in \mathbb{N}$ un número natural, $\Omega \subset \mathbb{R}^n$ un subconjunto de $\mathbb{R}^n$ y $f : \Omega \to \mathbb{R}$ una función real definida en Ω.*
Si existe el menor número $\alpha \in \mathbb{R}$ tal que $f(\boldsymbol{x}) \leq \alpha$ para todo $\boldsymbol{x} \in \Omega$, entonces α es el ***superemo*** *de f en Ω y se denota por $\sup_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}) = \alpha$.*
Si existe el mayor número $\beta \in \mathbb{R}$ tal que $f(\boldsymbol{x}) \geq \beta$ para todo $\boldsymbol{x} \in \Omega$, entonces β es el ***ínfimo*** *de f en Ω y se denota por $\inf_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}) = \beta$.*

Nota 18. *En la situación de la definición anterior, si $\boldsymbol{x^0} \in \Omega$ es un máximo global (resp. mínimo global) de f en Ω, entonces $f(\boldsymbol{x^0}) = \sup_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$ (resp. $f(\boldsymbol{x^0}) = \inf_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$).*

Las cantidades $\mathfrak{I}$ y $\mathfrak{S}$ para el programa (PC) son definidas por los siguientes ínfimo y supremo

$$\mathfrak{I} = \inf\{f(\boldsymbol{x}) : \boldsymbol{x} \in \Omega, h_1(\boldsymbol{x}) \leq 0, ..., h_m(\boldsymbol{x}) \leq 0\} \quad \text{y} \quad \mathfrak{S} = \sup_{\boldsymbol{\lambda}\geq 0}\left\{\inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})\right\}$$

Evidentemente, si $\boldsymbol{x^0} \in \mathbb{R}^n$ es un punto factible con $f(\boldsymbol{x^0}) = \mathfrak{I}$, entonces $\boldsymbol{x^0}$ es una solución de (PC) y si f no esta acotada inferiormente en la región factible, entonces $\mathfrak{I} = -\infty$ y (PC) no tiene solución.

Un vector $\boldsymbol{\lambda} \in \mathbb{R}^m$ con $\boldsymbol{\lambda} \geq \boldsymbol{0}$ es **factible** para (PCD) si $\inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) > -\infty$. El programa dual (PCD) es **consistente** si existe, al menos, un vector factible para (PCD). Un

vector factible $\boldsymbol{\lambda^0} \in \mathbb{R}^m$ para (PCD) es una **solución** para (PCD) si $g(\boldsymbol{\lambda^0}) = \sup_{\boldsymbol{\lambda}\geq 0} g(\boldsymbol{\lambda})$.

Existe una estrecha relación entre las magnitudes $\mathfrak{I}$ y $\mathfrak{S}$. La clave para encontrar tal vínculo está en el Teorema del punto de silla de Kuhn y Tucker, que ya apareció en el Capítulo 2 y enunciamos enseguida:

Teorema 16 (Punto de silla)**.** *Si (PC) es un programa convexo y superconsistente, entonces $\boldsymbol{x^0} \in \Omega$ es una solución de (PC) si, y sólo si, existe un vector $\boldsymbol{\lambda^0} \in \mathbb{R}^m$ tal que*

1. $\boldsymbol{\lambda^0} = (\lambda_1^0, ..., \lambda_m^0) \geq \mathbf{0}$
2. $\mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda}) \leq \mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda^0}) \leq \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda^0}) \quad \forall \boldsymbol{x} \in \Omega, \forall \boldsymbol{\lambda} \geq \mathbf{0}$
3. $\lambda_k^0 h_k(\boldsymbol{x^0}) = 0 \quad \forall k = 1, ..., m$

La demostración de este resultado puede verse en el capítulo 5 de [40].

Supóngase que $\boldsymbol{x^0} \in \mathbb{R}^n$ es una solución de (PC) y que existe un multiplicador $\boldsymbol{\lambda^0} \in \mathbb{R}^n$ de Karush-Kuhn-Tucker asociado a $\boldsymbol{x^0}$ cumpliendo las tres condiciones del teorema de arriba. Por un lado, la condición 2 del teorema anterior implica

$$\inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \leq \mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda}) \leq \mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda^0})$$

y consecuentemente

$$\sup_{\boldsymbol{\lambda}\geq 0} \left\{ \inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \right\} \leq \mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda^0})$$

y por otro lado, como $\mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda^0}) \leq \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda^0})$ para todo $\boldsymbol{x} \in \Omega$, se tiene que

$$\mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda^0}) \leq \inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda^0}) \leq \sup_{\boldsymbol{\lambda}\geq 0} \left\{ \inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \right\}$$

Se concluye entonces que $\mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda^0}) = \sup_{\boldsymbol{\lambda}\geq 0} \{\inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})\}$ pero obsérvese que, en virtud de la condición 3 del teorema anterior, también

$$\mathfrak{I} = f(\boldsymbol{x^0}) = f(\boldsymbol{x^0}) + \sum_{k=1}^{m} \lambda_k^0 h_k(\boldsymbol{x^0}) = \mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda^0})$$

Por lo tanto

$$\mathfrak{I} = \sup_{\boldsymbol{\lambda}\geq 0} \left\{ \inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \right\}$$

Estos cálculos muestran que si $\boldsymbol{x^0}$ es una solución de (PC) y que si $\boldsymbol{\lambda^0}$ es correspondiente multiplicador de Karush-Kuhn-Tucker, entonces $\boldsymbol{\lambda^0}$ es una solución de (PCD) y $\mathfrak{I} = \mathfrak{S}$. En particular, si un programa convexo y superconsistente tiene solución $\boldsymbol{x^0}$, entonces existe un multiplicador de Karush-Kuhn-Tucker $\boldsymbol{\lambda^0}$ que es solución del programa dual asociado. Así, los multiplicadores de Karush-Kuhn-Tucker del programa primal son soluciones del programa dual. Además, si $\boldsymbol{\lambda^0}$ es un multipicadore de Karush-Kuhn-Tucker para el programa (PC) y $\boldsymbol{x^0}$ es solución de (PC), entonces como

$$f(\boldsymbol{x^0}) = \mathfrak{I} = \mathcal{L}(\boldsymbol{x^0}, \boldsymbol{\lambda^0}) = \inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

se sigue que $\boldsymbol{x^0}$ puede ser buscado minimizando la función $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda^0})$ sobre Ω.

Así, al menos en teoría, ante un programa convexo (PC) se puede proceder de la siguiente manera:
Paso 1: Construir el programa dual (PCD) asociado al programa primal (PC).
Paso 2: Encontrar las soluciones $\boldsymbol{\lambda^0}$ de (PCD).
Paso 3: Encontrar las correspondientes soluciones de (PC) minimizando $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda^0})$ en Ω, donde $\mathcal{L}$ representa el lagrangiano asociado a (PC).

Obviamente, esta manera de proceder es útil solo cuando el programa (PCD) resulta mas fácil de resolver que el propio (PC) y, aunque esto no ocurre siempre, si que pasa muy a menudo. A continuación, se demuestra el *Teorema de dualidad* para problemas convexos.

Teorema 17 (dual para Programación convexa)**.** *Si $\boldsymbol{y} \in \Omega$ es un punto factible para el programa convexo primal (PC) y $\boldsymbol{\lambda} \in \mathbb{R}^m$ es un vector factible para el programa convexo dual (PCD), entonces $f(\boldsymbol{y}) \geq g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$.*
Como consecuencia, si (PC) y (PCD) son ambos programas consistentes, entonces $\mathfrak{I}$ y $\mathfrak{S}$ son ambos finitos y se da la desigualdad primal-dual, $\mathfrak{I} \geq \mathfrak{S}$.

Demostración. Por ser $\boldsymbol{y}$ un punto factible para (PC) y $\boldsymbol{\lambda}$ un vector factible para (PD), se tiene que

$$f(\boldsymbol{y}) \geq f(\boldsymbol{y}) + \sum_{k=1}^{m} \lambda_k h_k(\boldsymbol{y}) = \mathcal{L}(\boldsymbol{y}, \boldsymbol{\lambda})$$

puesto que $\lambda_k \geq 0$ y $h_k(\boldsymbol{y}) \leq 0$ para $k = 1, ..., m$. De aquí, se sigue que

$$f(\boldsymbol{y}) \geq g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

que es precisamente la primera afirmación. Esta desigualdad muestra también que

$$\mathfrak{S} = \sup_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \left\{ \inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \right\} \leq f(\boldsymbol{y})$$

y que

$$\mathfrak{I} = \inf \{ f(\boldsymbol{x}) : \boldsymbol{x} \in \Omega, h_1(\boldsymbol{x}) \leq 0, ..., h_m(\boldsymbol{x}) \leq 0 \} \geq g(\boldsymbol{\lambda})$$

siempre y cuando $\boldsymbol{y}$ y $\boldsymbol{\lambda}$ sean factibles para (PC) y (PCD), respectivamente. Esto prueba que si (PC) y (PCD) son ambos programas consistentes, entonces $\mathfrak{I}$ y $\mathfrak{S}$ son finitos y $\mathfrak{I} \geq \mathfrak{S}$. □

Corolario 9. *Si $\boldsymbol{x^0} \in \Omega$ es un punto factible para un programa convexo (PC) y $\boldsymbol{\lambda^0} \in \mathbb{R}^m$ es un vector factible para el programa convexo dual (PCD) tales que $f(\boldsymbol{x^0}) = g(\boldsymbol{\lambda^0})$, entonces $\boldsymbol{x^0}$ es una solución de (PC) y $\boldsymbol{\lambda^0}$ es una solución de (PCD).*

Demostración. Según la Desigualdad primal-dual y las definiciones de $\mathfrak{I}$ y $\mathfrak{S}$, es sabido que $f(\boldsymbol{x^0}) \geq \mathfrak{I} \geq \mathfrak{S} \geq g(\boldsymbol{\lambda})$, pero por hipótesis se da la igualdad en todas las desigualdades anteriores por lo que $f(\boldsymbol{x^0}) = \mathfrak{I}$ y $g(\boldsymbol{\lambda^0}) = \mathfrak{S}$, lo que implica que $\boldsymbol{x^0}$ y $\boldsymbol{\lambda^0}$ son soluciones de (PC) y (PCD), respectivamente. □

Finalmente, se ilustra esta técnica para tres casos concretos: la Programación lineal, geométrica y cuadrática. Sólo problemas de minimización serán considerados.

3.4.1. Programación lineal

La **Programación lineal** es una de las áreas más importantes en el campo de la optimización restringida. Su importancia surge, por una parte, por el amplio rango de problemas de matemática aplicada que se resuelven con programas lineales (por ejemplo, el *problema de la dieta* o *problemas de marketing* entre otros, [40]) y, por otra parte, por la gran efectividad de los métodos matemáticos desarrollados para resolver tales programas (por ejemplo, el *método simplex* entre otros, [40]). En lo que sigue, se verá que la Programación lineal es un caso particular de la Programación convexa y se obtendrá el *Teorema de la Programación lineal dual*, como consecuencia directa del teorema 17.

Dados $n, m \in \mathbb{N}$ dos números naturales, un **programa lineal restringido** es un problema de optimización restringida donde, como su propio nombre indica, la función objetivo f y las funciones restricciones $g_1, ..., g_m$ son funciones lineales definidas en todo $\mathbb{R}^n$. Si $A = (a_{ij}) \in \mathfrak{M}_{m\times n}(\mathbb{R})$ es una matriz $m \times n$ con coeficientes reales y $\boldsymbol{b} = (b_1, ..., b_n) \in \mathbb{R}^n$ y $\boldsymbol{c} = (c_1, ..., c_m) \in \mathbb{R}^m$ son dos vectores, entonces un programa lineal viene descrito como

$$\left.\begin{array}{l} \text{minimizar } b_1x_1 + b_2x_2 + \cdots + b_nx_n \\ \text{sujeto a} \\ \qquad a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \geq c_1 \\ \qquad a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \geq c_2 \\ \qquad\qquad \vdots \\ \qquad a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \geq c_m \\ \text{donde } x_1 \geq 0, x_2 \geq 0, ..., x_n \geq 0 \end{array}\right\} (PL)$$

Si aceptamos que $\boldsymbol{x} = (x_1, ..., x_n) \geq \boldsymbol{0}$ significa $x_1 \geq 0, ..., x_n \geq 0$ y que $A\boldsymbol{x} \geq \boldsymbol{c}$ significa $\langle \boldsymbol{a}^{(k)}, \boldsymbol{x}\rangle \geq c_k$ para todo $k = 1, ..., m$ donde $\boldsymbol{a}^{(k)}$ representa la columna k-ésima de la matriz A, entonces podemos reformular el programa (PL) de la siguiente manera

$$\left.\begin{array}{l} \text{minimizar } f(\boldsymbol{x}) = \langle \boldsymbol{b}, \boldsymbol{x}\rangle \\ \text{sujeto a} \\ \qquad g_1(\boldsymbol{x}) = c_1 - \langle \boldsymbol{a}^{(1)}, \boldsymbol{x}\rangle \leq 0 \\ \qquad\qquad \vdots \\ \qquad g_m(\boldsymbol{x}) = c_m - \langle \boldsymbol{a}^{(m)}, \boldsymbol{x}\rangle \leq 0 \\ \text{donde } \boldsymbol{x} \in \Omega = (\mathbb{R}_0^+)^n \end{array}\right\} (PL)$$

Obsérvese que las funciones f y g_k con $k = 1, ..., m$ son lineales y, por tanto, convexas (véase el ejemplo 16) y evidentemente también lo es Ω, luego (PL) es un programa convexo. Esto prueba que todo programa lineal es también convexo y lo escribiremos en forma compacta como

$$\left.\begin{array}{c} \text{minimizar } f(\boldsymbol{x}) = \langle \boldsymbol{b}, \boldsymbol{x}\rangle \\ \text{sujeto a } A\boldsymbol{x} \geq \boldsymbol{c} \text{ donde } \boldsymbol{x} \geq \boldsymbol{0} \end{array}\right\} (PL)$$

En ese caso, el lagrangiano de (PL) está dado por

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \langle \boldsymbol{b}, \boldsymbol{x}\rangle + \sum_{k=1}^{m} \lambda_k(c_k - \langle \boldsymbol{a}^{(k)}, \boldsymbol{x}\rangle) = \langle \boldsymbol{b} - \sum_{k=1}^{m} \lambda_k \boldsymbol{a}^{(k)}, \boldsymbol{x}\rangle + \sum_{k=1}^{m} \lambda_k c_k = \langle \boldsymbol{b} - A^t\boldsymbol{\lambda}, \boldsymbol{x}\rangle + \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle$$

para todo $(\boldsymbol{x}, \boldsymbol{\lambda}) \in \Omega \times \mathbb{R}^n$.

Para construir el programa dual (PLD) asociado a (PL), se comienza por identificar la región factible para (PLD), esto es, el conjunto de puntos $\boldsymbol{\lambda} \geq \mathbf{0}$ para los que $\inf_{\boldsymbol{x}\in\Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) > -\infty$. Por un lado, nótese que si existe algún $j_0 \in \{1, ..., n\}$ tal que $b_{j_0} - (A^T\boldsymbol{\lambda})_{j_0} < 0$ y si definimos el vector $\boldsymbol{x}^{(t)}$ para cada $t < 0$ como

$$x_i^{(t)} = \begin{cases} t(b_{j_0} - (A^T\boldsymbol{\lambda})_{j_0}) & \text{si } i = j_0 \\ 0 & \text{si } i \neq j_0 \end{cases}$$

entonces $\boldsymbol{x}^{(t)} \geq \mathbf{0}$ para $t < 0$ y $\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}) = t(b_{j_0} - (A^T\boldsymbol{\lambda})_{j_0})^2 + \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle$, de donde se desprende que $\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}) \to -\infty$ cuando $t \to -\infty$. Asi que, $\boldsymbol{\lambda}$ no es un punto factible para (PLD). Por otro lado, nótese que si $\boldsymbol{b} - A^T\boldsymbol{\lambda} \geq \mathbf{0}$, entonces

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \langle \boldsymbol{b} - A^T\boldsymbol{\lambda}, \boldsymbol{x}\rangle + \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle \geq \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle$$

para cada $\boldsymbol{x} \geq \mathbf{0}$, luego la región factible de (PLD) es $\{\boldsymbol{\lambda} \in \mathbb{R}^m : \boldsymbol{\lambda} \geq \mathbf{0}, A^T\boldsymbol{\lambda} \leq \boldsymbol{b}\}$.

Finalmente, obsérvese que $g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}\geq\mathbf{0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle$ porque $\mathcal{L}(\mathbf{0}, \boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle$ y $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \geq \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle$ para todo $\boldsymbol{x} \geq \mathbf{0}$. Consecuentemente, el programa dual de (PL) puede formularse como

$$\left.\begin{array}{c} \text{maximizar } g(\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle \\ \text{sujeto a } A^T\boldsymbol{\lambda} \leq \boldsymbol{b} \text{ donde } \boldsymbol{\lambda} \geq \mathbf{0} \end{array}\right\} (PLD)$$

Por consiguiente, el teorema 17 da lugar a la siguiente versión práctica de Teorema dual para el caso de la Programación lineal.

Teorema 18 (Teorema dual para la Programación lineal)**.** *Si el programa (PL) es superconsistente y $\boldsymbol{x^0} \in \Omega$ es solución de (PL), entonces el programa (PLD) tiene solución $\boldsymbol{\lambda^0} \in \mathbb{R}^m$ y $\langle \boldsymbol{b}, \boldsymbol{x^0}\rangle = f(\boldsymbol{x^0}) = g(\boldsymbol{\lambda^0}) = \langle \boldsymbol{c}, \boldsymbol{\lambda^0}\rangle$.*
Además, si $\boldsymbol{x} \in \Omega$ es cualquier punto factible para (PL) y $\boldsymbol{\lambda} \in \mathbb{R}^m$ es cualquier punto factible para (PLD), entonces $\langle \boldsymbol{b}, \boldsymbol{x}\rangle \geq \langle \boldsymbol{\lambda}, \boldsymbol{c}\rangle$.

Ilustramos este teorema en un ejemplo concreto.

Ejemplo 24. *Considérese el programa lineal*

$$\left.\begin{array}{c} \textit{minimizar } f(x, y, z, t) = 4x + 15y + 12z + 2t \\ \textit{sujeto a} \begin{cases} 2y + 3z + t \geq 1 \\ x + 3y + z - t \geq 1 \\ x, y, z, t \geq 0 \end{cases} \end{array}\right\}$$

En notación matricial sería

$$\left.\begin{array}{c} \textit{minimizar } f(\boldsymbol{x}) = \langle \boldsymbol{b}, \boldsymbol{x}\rangle \\ \textit{sujeto a } A\boldsymbol{x} \geq \boldsymbol{c}, \boldsymbol{x} \geq \mathbf{0} \end{array}\right\} (PL)$$

donde $\boldsymbol{b} = (4, 15, 12, 2)$, $\boldsymbol{c} = (1, 1)$ y $A = \begin{pmatrix} 0 & 2 & 3 & 1 \\ 1 & 3 & 1 & -1 \end{pmatrix}$.

El programa dual asociado es

$$\left.\begin{array}{l} \textit{maximizar } g(\boldsymbol{\lambda}) = \lambda_1 + \lambda_2 \\ \textit{sujeto a } \begin{cases} \lambda_2 \leq 4 \\ 2\lambda_1 + 3\lambda_2 \leq 15 \\ 3\lambda_1 + \lambda_2 \leq 12 \\ \lambda_1 - \lambda_2 \leq 2 \\ \lambda_1, \lambda_2 \geq 0 \end{cases} \end{array}\right\} (PLD)$$

Este último sistema en λ_1 y λ_2 puede resolverse gráficamente (véase [40]) y no es dificil ver que la solución es $\lambda_1^0 = \lambda_2^0 = 3$ y, por tanto, el máximo de la función g vale

$$g(\lambda_1^0, \lambda_2^0) = 3 + 3 = 6 = \mathfrak{S}$$

La condición de holgura dice que

$$\begin{cases} 3(1 - 2y - 3z - t) = 0 \\ 3(1 - x - 3y - z + t) = 0 \end{cases}$$

y el teorema 17 nos dice que

$$4x + 15y + 12z + 2t = 6$$

lo que conduce al sistema de tres ecuaciones con cuatro incógnitas siguiente

$$\begin{cases} 2y + 3z + t = 1 \\ x + 3y + z - t = 1 \\ 4x + 15y + 12z + 2t = 6 \end{cases}$$

que tiene por solución

$$x = -2t, \quad y = \frac{2 + 10t}{7}, \quad z = \frac{1 - 9t}{7}$$

pero toda solución de (PL) debe tener componentes no negativas, lo que obliga a que sea $t = 0$ y con ello a que

$$x_0 = 0, \quad y_0 = \frac{2}{7}, \quad z_0 = \frac{1}{7}, \quad t_0 = 0$$

sea la única solución de (PL).

3.4.2. Programación geométrica

A continuación, se presenta como la desigualdad aritmético-geométrica (ec. 3.8) resulta adecuada para resolver una gran clase de problemas de optimización no lineales. Antes de identificar dicha clase de problemas, se exponen algunos ejemplos de aplicación de la desigualdad en el campo de la optimización no restringida.

Ejemplo 25 (Caja rectangular abierta de área fija dada con mayor volumen)**.** *Sean x, y, z las medidas de una caja rectangular abierta (largo, ancho y altura). Su volumen es $V = xyz$ y su área dada es $S_0 = xy + 2xz + 2yz$, luego el objetivo es resolver el siguiente problema*

$$\left.\begin{array}{c} \textit{maximizar } V(x,y,z) = xyz \\ \textit{sujeto a } xy + 2xz + 2yz = S_0 \end{array}\right\}$$

En virtud de la desigualdad aritmético-geométrica,

$$\begin{aligned} S_0 &= xy + 2xz + 2yz = 3\frac{xy + 2xz + 2yz}{3} \geq \\ &\geq 3[(xy)^{1/3}(2xz)^{1/3}(2yz)^{1/3}] = 3 \cdot 4^{1/3}(x^2y^2z^2)^{1/3} = 3 \cdot 4^{1/3}V^{2/3} \end{aligned}$$

Así, V es máximo cuando se da la igualdad en la Desigualdad aritmético-geométrica, esto es, V es máximo cuando

$$xy = 2xz = 2yz = \frac{S_0}{3}$$

Un cálculo breve muestra que

$$x_0 = y_0 = \sqrt{\frac{S_0}{3}} \quad y \quad z_0 = \frac{1}{2}\sqrt{\frac{S_0}{3}}$$

y consecuentemente, el volumen máximo de la caja es

$$V_0 = x_0y_0z_0 = \frac{S_0^{3/2}}{2 \cdot 3^{3/2}}$$

Ejemplo 26 (Caja rectangular abierta de volumen fijo dado con menor área)**.** *Sean x, y, z las medidas de una caja rectangular abierta (largo, ancho y altura). Su área es $S = xy + 2xz + 2yz$ y su volumen dado es $V_0 = xyz$, luego el objetivo es resolver el siguiente problema*

$$\left.\begin{array}{c} \textit{minimizar } S(x,y,z) = xy + 2xz + 2yz \\ \textit{sujeto a } xyz = V_0 \end{array}\right\}$$

Procediendo como en el ejemplo anterior, en virtud de la desigualdad aritmético-geométrica,

$$S = 3\frac{xy + 2xz + 2yz}{3} \geq 3 \cdot 4^{1/3}V_0^{2/3}$$

se sigue que S es mínima cuando hay igualdad en la desigualdad aritmético-geométrica, esto es, S es mínima cuando

$$xy = 2xz = 2yz = 4^{1/3}V_0^{2/3}$$

Un cálculo breve muestra que

$$x_0 = y_0 = 0 = 4^{1/6}V_0^{1/3} \quad y \quad z_0 = \frac{4^{1/6}V_0^{1/3}}{2}$$

y consecuentemente, el área mínima de la caja es

$$S_0 = x_0y_0 + 2x_0z_0 + 2y_0z_0 = 3 \cdot 4^{1/3}V_0^{2/3}$$

La pareja de los dos ejemplos anteriores proporcionan un primer ejemplo de *programas geométricos duales*. En ambos ejemplos, se tratan esencialmente con las mismas funciones salvo que la función objetivo de un ejemplo es la restricción del otro y que un ejemplo es de minimización y el otro de maximización. El próximo ejemplo provee una visión mas profunda de dualidad.

Ejemplo 27. *Considérese los siguientes programas*

$$\left.\begin{array}{c} \textit{maximizar } f(x,y,z) = xy^2z \\ \textit{sujeto a } g_1(x,y,z) = x+y+z^2 = \kappa \\ \textit{donde } \kappa > 0 \textit{ es fijo y } x,y,z \in \mathbb{R}^+ \end{array}\right\} (P) \qquad \left.\begin{array}{c} \textit{minimizar } h(x,y,z) = x+y+z^2 \\ \textit{sujeto a } g_2(x,y,z) = xy^2z = \xi \\ \textit{donde } \xi > 0 \textit{ es fijo y } x,y,z \in \mathbb{R}^+ \end{array}\right\} (D)$$

Los programas (P) y (D) son programas duales: (P) es el programa dual de (D) y, viceversa, (D) es el programa dual de (P).

Para resolver ambos programas, se necesita expresar $x+y+z^2$ de forma que la aplicación de la desigualdad aritmético-geométrica proporcione una cota inferior de xy^2z. Esto puede conseguirse como sigue

$$\begin{aligned} x+y+z^2 &= \frac{x}{2}+\frac{x}{2}+\frac{y}{4}+\frac{y}{4}+\frac{y}{4}+\frac{y}{4}+z^2 = \\ &= 7\frac{\frac{x}{2}+\frac{x}{2}+\frac{y}{4}+\frac{y}{4}+\frac{y}{4}+\frac{y}{4}+z^2}{7} \geq \\ &\geq 7\left(\frac{1}{2}\right)^{2/7}\left(\frac{1}{4}\right)^{4/7}(xy^2z)^{2/7} \end{aligned}$$

La igualdad se da precisamente cuando

$$\frac{x}{2} = \frac{y}{4} = z^2$$

Para maximizar f sujeto a la restricción dada por g_1, se toman los valores óptimos x_0, y_0 y z_0 que fuerzan la igualdad en la desigualdad de las medias y se igualan, en este caso, a $\kappa/7$:

$$\frac{x}{2} = \frac{y}{4} = z^2 = \frac{\kappa}{7} \Rightarrow \begin{cases} x_0 = \frac{2\kappa}{7} \\ y_0 = \frac{4\kappa}{7} \\ z_0 = \sqrt{\frac{\kappa}{7}} \end{cases}$$

Para minimizar h sujeto a la restricción dada por g_2, se toman los valores óptimos x_0, y_0 y z_0 que fuerzan la igualdad en la desigualdad de las medias y se igualan, en este caso, a $\left(\frac{1}{2}\right)^{2/7}\left(\frac{1}{4}\right)^{4/7}\xi^{2/7}$:

$$\frac{x}{2} = \frac{y}{4} = z^2 = \left(\frac{1}{2}\right)^{2/7}\left(\frac{1}{4}\right)^{4/7}\xi^{2/7} \Rightarrow \begin{cases} x_0 = 2\left(\frac{1}{2}\right)^{2/7}\left(\frac{1}{4}\right)^{4/7}\xi^{2/7} \\ y_0 = 2x_0 \\ z_0 = \sqrt{\frac{x_0}{2}} \end{cases}$$

La desigualdad aritmético-geométrica también sirve para resolver algunos problemas de minimización no restringida, como es el caso del próximo ejemplo.

Ejemplo 28 (Encontrar el mínimo de $f(x,y) = 4x + \dfrac{x}{y^2} + 4\dfrac{y}{x}$ para $x, y \in \mathbb{R}^+$). *De nuevo, este problema puede resolverse usando la desigualdad aritmético-geométrica de la siguiente manera*

$$f(x,y) = 4\frac{4x + \frac{x}{y^2} + 2\frac{y}{x} + 2\frac{y}{x}}{4} \geq 4(4^{1/4})(2^{2/4})\left(\frac{x^2y^2}{y^2x^2}\right)^{1/4} = 8$$

Si se fuerza la igualdad en la desigualdad anterior, se obtiene que los valores x_0 e y_0 que minimizan f vienen dados por

$$4x_0 = \frac{x_0}{y_0^2} = \frac{2y_0}{x_0} = 2 \Leftrightarrow x_0 = y_0 = \frac{1}{2}$$

Consecuentemente, el valor mínimo de f en $\mathbb{R}^+ \times \mathbb{R}^+$ es 8 y se alcanza en $(1/2, 1/2)$.

Nota 19. *Animo al lector a intentar resolver el problema anterior usando las técnicas expuestas en el Capítulo 1.*

Todos los ejemplos discutidos hasta ahora parecen indicar que la desigualdad arimético-geométrica puede ser una herramienta útil para resolver ciertos problemas de optimización. Estos problemas son los nombrados *programas geométricos*. Como se ha visto con los ejemplos anteriores, a menudo es posible resolver tales problemas (programas geométricos) usando solamente la desigualdad aritmético-geométrica. Sin embargo, algunos de estos problemas son muy complicados para proceder de esta manera. Aquí veremos un procedimiento que nos rescatará de estas situaciones. También, el lector habrá podido alertar que quizás no siempre se pueda forzar la igualdad en la desigualdad de las medias como se ha exigido en los ejemplos, pero una vez desarrollado esta teoría se verá que efectivamente sí se puede y todo lo anterior tiene sentido.

Definición 22 (Posinomio). *Dados n,m $\in \mathbb{N}$ dos número naturales, un* ***posinomio*** *en m-variables $\boldsymbol{t} = (t_1, ..., t_m) \in (\mathbb{R}^+)^m$ es una función $\mathfrak{p} : \mathbb{R}^m \to \mathbb{R}$ definida por la siguiente combinación lineal*

$$\mathfrak{p}(\boldsymbol{t}) = \sum_{i=1}^{n} \alpha_i \prod_{j=1}^{m} (t_j)^{\beta_{ij}} \qquad \forall \boldsymbol{t} \in (\mathbb{R}^+)^m$$

para ciertos coeficientes $\{\alpha_i \in \mathbb{R}^+ : i = 1, ..., n\}$ y ciertos exponentes $\{\beta_{ij} \in \mathbb{R} = 1, ..., n, j = 1, ..., m\}$

El objetivo de la **Programación geométrica** no restringida es resolver el siguiente programa

$$\left.\begin{array}{c} \text{minimizar } \mathfrak{p}(\boldsymbol{t}) \text{ donde} \\ \boldsymbol{t} = (t_1, ..., t_m) \in \mathbb{R}^m \text{ y } t_1 > 0, ..., t_m > 0 \end{array}\right\} \quad (PG)$$

Una **solución** de (PG) es un mínimo global de $\mathfrak{p}$ en la región factible de vectores $\boldsymbol{t} = (t_1, ..., t_m)$ con componentes positivas. Antes de nada, obsérvese que $\mathfrak{p}$ puede reescribirse como

$$\mathfrak{p}(\boldsymbol{t}) = \sum_{i=1}^{n} \delta_i \left(\frac{\alpha_i \prod_{j=1}^{m} t_j^{\beta_{ij}}}{\delta_i}\right)$$

donde cada δ_i se supone positivo $\delta_i > 0$ (*condición de positividad*). Si añadimos la restricción $\delta_1 + \cdots + \delta_m = 1$ (*condición de normalidad*) entonces podemos aplicar la desigualdad aritmético-geométrica en esta nueva expresión de $\mathfrak{p}$ y obtener

$$\mathfrak{p}(\boldsymbol{t}) \geq \prod_{i=1}^{n} \left(\frac{\alpha_i \prod_{j=1}^{m} t_j^{\beta_{ij}}}{\delta_i} \right)^{\delta_i} = \prod_{i=1}^{n} \left(\frac{\alpha_i}{\delta_i} \right)^{\delta_i} \left(\prod_{i=1}^{n} \prod_{j=1}^{m} t_j^{\beta_{ij}\delta_i} \right) = \prod_{i=1}^{n} \left(\frac{\alpha_i}{\delta_i} \right)^{\delta_i} \prod_{j=1}^{m} t_j^{\sum_{i=1}^{n} \beta_{ij}\delta_i}$$

Por tanto, si imponemos la condición adicional $\sum_{i=1}^{n} \beta_{ij}\delta_i = 0$ para todo $j = 1, ..., m$ (*condición de ortogonalidad*), la desigualdad de arriba afirma

$$\mathfrak{p}(\boldsymbol{t}) \geq \prod_{i=1}^{n} \left(\frac{\alpha_i}{\delta_i} \right)^{\delta_i}$$

Así, definiendo $\mathfrak{q}(\boldsymbol{\delta}) = \prod_{i=1}^{n} \left(\frac{\alpha_i}{\delta_i} \right)^{\delta_i}$, queda probada la Desigualdad primal-dual:

$$\mathfrak{p}(\boldsymbol{t}) \geq \mathfrak{q}(\delta) \tag{3.9}$$

para todo $\boldsymbol{t} \in \mathbb{R}^n$ con componentes positivas y todo $\boldsymbol{\delta} \in \mathbb{R}^n$ satisfaciendo las condiciones de positividad, normalidad y ortogonalidad. Esto conlleva a considerar el *programa geométrico dual* de (PG)

$$\left.\begin{array}{l} \text{maximizar } \mathfrak{q}(\delta) = \displaystyle\prod_{i=1}^{n} \left(\frac{\alpha_i}{\delta_i} \right)^{\delta_i} \\ \text{sujeto a } \begin{cases} \delta_1 > 0, ..., \delta_n > 0 \\ \delta_1 + \cdots + \delta_n = 1 \\ \sum_{i=1}^{n} \beta_{ij}\delta_i = 0 \quad \forall j = 1, ..., m \end{cases} \end{array}\right\} \quad (PGD)$$

Un vector $\boldsymbol{\delta} \in \mathbb{R}^n$ es **factible** para (PGD) si satisface las condiciones de positividad, normalidad y ortogonalidad. El programa dual (PGD) es **consistente** si el conjunto de vectores factibles de (PGD) es no vacío. Finalmente, una **solución** de (PGD) es un vector $\boldsymbol{\delta^0} \in \mathbb{R}^n$ que maximiza $\mathfrak{q}(\boldsymbol{\delta})$ en el conjunto de puntos factibles de (PGD).

Nótese que si $\boldsymbol{t^0} \in (\mathbb{R}^m)^+$ es solución de (PG) y $\boldsymbol{\delta^0} \in \mathbb{R}^n$ es solución de (PGD), en virtud de la desigualdad 3.9, se tiene que $\mathfrak{p}(\boldsymbol{t^0}) \geq \mathfrak{q}(\boldsymbol{\delta^0})$. Vamos a ver que, en realidad, se da la igualdad $\mathfrak{p}(\boldsymbol{t^0}) = \mathfrak{q}(\boldsymbol{\delta^0})$ y que dicha igualdad establece un proceso directo para computar soluciones de (PG) y (PGD) siempre y cuando estos programas tengan solución.

Teorema 19 (Programación geométrica no restringida)**.** *Si $\boldsymbol{t^0} = (t_1^0, ..., t_m^0) \in (\mathbb{R}^m)^+$ es una solución de (PG), entonces el correspondiente programa dual (PGD) es consistente. De hecho, el vector $\boldsymbol{\delta^0} = (\delta_1^0, ..., \delta_n^0) \in \mathbb{R}^n$ definido por*

$$\delta_i^0 = \frac{\tau_i(\boldsymbol{t^0})}{\mathfrak{p}(\boldsymbol{t^0})} \qquad \forall i = 1, ..., n$$

donde $\tau_i(\boldsymbol{t}) = \alpha_i t_1^{\beta_{i1}} \cdots t_m^{\beta_{im}}$ es el i-ésimo término de $\mathfrak{p}$, es una solución de (PGD) y se da la igualdad en la Desigualdad primal-dual, $\mathfrak{p}(\boldsymbol{t^0}) = \mathfrak{q}(\boldsymbol{\delta^0})$.

Demostración. Sea $i \in \{1, ..., n\}$ y $\tau_i(\boldsymbol{t}) = \alpha_i t_1^{\beta_{i1}} \cdots t_m^{\beta_{im}}$ el i-ésimo término de $\mathfrak{p}$. Dado $j \in \{1, ..., m\}$, se tiene que

$$t_j \frac{\partial \tau_i(\boldsymbol{t})}{\partial t_j} = \beta_{ij} \tau_i(\boldsymbol{t})$$

Como $\boldsymbol{t^0}$ es un mínimo de $\mathfrak{p}$, se sigue que

$$0 = \frac{\partial \mathfrak{p}(\boldsymbol{t^0})}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial \tau_i(\boldsymbol{t^0})}{\partial t_j} \quad \forall j = 1, ..., m$$

Pero entonces, por la primera observación hecha al empezar la prueba, es claro que

$$0 = \sum_{i=1}^{n} \beta_{ij} \tau_i(\boldsymbol{t^0}) \quad \forall j = 1, ..., m$$

Finalmente, como $\mathfrak{p}(\boldsymbol{t^0}) > 0$, se puede dividir ambos miembros de esta última ecuación por $\mathfrak{p}(\boldsymbol{t^0})$ y obtener

$$0 = \sum_{i=1}^{n} \beta_{ij} \frac{\tau_i(\boldsymbol{t^0})}{\mathfrak{p}(\boldsymbol{t^0})} \quad \forall j = 1, ..., m$$

Consecuentemente, si se toma

$$\delta_i^0 = \frac{\tau_i(\boldsymbol{t^0})}{\mathfrak{p}(\boldsymbol{t^0})} \quad \forall i = 1, ..., n$$

entonces $\boldsymbol{\delta^0}$ satisface la condición de ortogonalidad para el programa dual (PGD). Además, $\delta_i^0 > 0$ para $i = 1, ..., n$ asi que la condición de positividad se verifica también. Por último,

$$\sum_{i=1}^{n} \delta_i^0 = \sum_{i=1}^{n} \frac{\tau_i(\boldsymbol{t^0})}{\mathfrak{p}(\boldsymbol{t^0})} = \frac{\mathfrak{p}(\boldsymbol{t^0})}{\mathfrak{p}(\boldsymbol{t^0})} = 1$$

luego se cumple la condición de normalidad. Se concluye, por tanto, que $\boldsymbol{\delta^0}$ es un vector factible para (PGD), así que (PGD) es consistente y también se da la igualdad en la Desigualdad primal-dual:

$$\begin{aligned}
\mathfrak{p}(\boldsymbol{t^0}) &= \mathfrak{p}(\boldsymbol{t^0})^{\delta_1^0 + \cdots + \delta_n^0} = (\mathfrak{p}(\boldsymbol{t^0}))^{\delta_1^0} + \cdots (\mathfrak{p}(\boldsymbol{t^0}))^{\delta_n^0} = \\
&= \left(\frac{\tau_1(\boldsymbol{t^0})}{\delta_1^0}\right)^{\delta_1^0} \cdots \left(\frac{\tau_n(\boldsymbol{t^0})}{\delta_n^0}\right)^{\delta_n^0} = \\
&= \left(\frac{\alpha_1}{\delta_1^0}\right)^{\delta_1^0} \cdots \left(\frac{\alpha_n}{\delta_n^0}\right)^{\delta_n^0} t_1^{\sum_{i=1}^n \beta_{i1}\delta_i^0} \cdots t_m^{\sum_{i=1}^n \beta_{im}\delta_i^0} = \\
&= \left(\frac{\alpha_1}{\delta_1^0}\right)^{\delta_1^0} \cdots \left(\frac{\alpha_n}{\delta_n^0}\right)^{\delta_n^0} = \mathfrak{q}(\boldsymbol{\delta^0})
\end{aligned}$$

□

Procedimiento para la Programación geométrica (no restringida). Dado un programa geométrico

$$\left.\begin{array}{c}\text{minimizar } \mathfrak{p}(t) = \sum_{i=1}^{n} \tau_i(\boldsymbol{t}) \text{ donde} \\ \boldsymbol{t} = (t_1, ..., t_m) \in (\mathbb{R}^m)^+ \text{ y } \tau_i(\boldsymbol{t}) = \alpha_i \prod_{j=1}^{m} (t_j)^{\beta_{ij}}\end{array}\right\} \quad (PG)$$

se procede como sigue:

PASO 1. Calcular la región factible del programa geométrico dual, esto es, el conjunto de vectores $\boldsymbol{\delta} = (\delta_1, ..., \delta_n) \in \mathbb{R}^n$ tales que

$$\delta_1 > 0, ..., \delta_n > 0 \qquad \text{(condición de positividad)}$$

$$\sum_{i=1}^{n} \delta_i = 1 \qquad \text{(condición de normalidad)}$$

$$\sum_{i=1}^{n} \beta_{ij}\delta_i = 0 \quad \forall j = 1, ..., m \qquad \text{(condición de ortogonalidad)}$$

PASO 2. Si la región factible de (PGD)

- es vacía, entonces (PG) no tiene solución.
- consiste únicamente de un vector $\boldsymbol{\delta^0}$, entonces $\boldsymbol{\delta^0}$ es una solución de (PGD) y se procede directamente al paso 4.
- consiste de mas de un vector, entonces se procede al paso 3.

PASO 3. Resolver (PGD).

PASO 4. Dada una solución $\boldsymbol{\delta^0}$ de (PGD), una solución $\boldsymbol{t^0}$ de (PG) se obtiene resolviendo las siguientes ecuaciones para $t_1^0, ..., t_m^0$

$$\delta_i^0 = \frac{\tau_i(\boldsymbol{t^0})}{\mathfrak{q}(\boldsymbol{\delta^0})} \quad \forall i = 1, ..., n$$

El valor mínimo $\mathfrak{p}(\boldsymbol{t^0})$ de $\mathfrak{p}$ es igual al valor máximo $\mathfrak{q}(\boldsymbol{\delta^0})$ de $\mathfrak{q}$.

Conviene hacer un par de aclaraciones y comentarios.

1. Para calcular la región factible de (PGD), primero se resuelve el sistema de ecuaciones lineales compuesto por las condiciones de normalidad y ortogonalidad, y después se impone la condición de positividad a las resultados obtenidos.

2. El hecho de que si la región factible de (PGD) es vacía, entonces (PG) no tiene solución se sigue del teorema 19. Si (PG) tiene solución, entonces (PGD) es consistente, es decir, tiene región factible no vacía.

3. La solución del sistema de ecuaciones del paso 4 parece complicado porque las ecuaciones no son lineales en las variables $t_1^0, ..., t_m^0$. Sin embargo, en [40] puede verse como se simplifica el cálculo tomando logaritmos.

Ejemplo 29 (Programación geométrica no restringida)**.** *Dados* $a, b, c, d, K \in \mathbb{R}^+$*, considérese el siguiente programa geométrico*

$$\left.\begin{array}{c} \textit{minimizar } \mathfrak{p}(\boldsymbol{t}) = \dfrac{aK}{t_1 t_2 t_3} + 2bt_2t_3 + 2ct_1t_3 + dt_1t_2 \\ \textit{donde } t_1 > 0, t_2 > 0, t_3 > 0, t_4 > 0 \end{array}\right\}$$

El correspondiente programa dual es

$$\left.\begin{array}{c} \textit{maximizar } \mathfrak{q}(\boldsymbol{\delta}) = \left(\dfrac{aK}{\delta_1}\right)^{1/\delta_1} \left(\dfrac{2b}{\delta_2}\right)^{1/\delta_2} \left(\dfrac{2c}{\delta_3}\right)^{1/\delta_3} \left(\dfrac{2d}{\delta_4}\right)^{1/\delta_4} \\ \textit{sujeto a } \begin{cases} \delta_1, \delta_2, \delta_3, \delta_4 > 0 \\ \delta_1 + \delta_2 + \delta_3 + \delta_4 = 1 \\ -\delta_1 + \delta_3 + \delta_4 = 0 \\ -\delta_1 + \delta_2 + \delta_4 = 0 \\ -\delta_1 + \delta_2 + \delta_3 = 0 \end{cases} \end{array}\right\}$$

En este caso, la región factible del programa dual consta solamente de un vector, a saber, $\boldsymbol{\delta^0} = \left(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$*. Por el paso 2, se tiene que la solución del programa primario se calcula resolviendo las siguiente ecuaciones para* t_1^0, t_2^0 *y* t_3^0

$$\begin{cases} \frac{aK}{t_1^0 t_2^0 t_3^0} = \delta_1^0 \mathfrak{q}(\boldsymbol{\delta^0}) = \frac{2}{5}\mathfrak{q}(\boldsymbol{\delta^0}) \\ 2bt_2^0 t_3^0 = \delta_2^0 \mathfrak{q}(\boldsymbol{\delta^0}) = \frac{1}{5}\mathfrak{q}(\boldsymbol{\delta^0}) \\ 2ct_1^0 t_3^0 = \delta_3^0 \mathfrak{q}(\boldsymbol{\delta^0}) = \frac{1}{5}\mathfrak{q}(\boldsymbol{\delta^0}) \\ dt_1^0 t_2^0 = \delta_4^0 \mathfrak{q}(\boldsymbol{\delta^0}) = \frac{1}{5}\mathfrak{q}(\boldsymbol{\delta^0}) \end{cases}$$

Se omiten los detalles de cálculo para resolver este sistema.

Ejemplo 30 (Programación geométrica no restringida)**.** *Considérese el siguiente programa geométrico*

$$\left.\begin{array}{c} \textit{minimizar } \mathfrak{p}(\boldsymbol{t}) = \dfrac{2}{t_1 t_2} + t_1t_2 + t_1 \\ \textit{donde } t_1 > 0, t_2 > 0 \end{array}\right\}$$

El correspondiente programa dual es

$$\left.\begin{array}{c} \textit{maximizar } \mathfrak{q}(\boldsymbol{\delta}) = \left(\dfrac{2}{\delta_1}\right)^{1/\delta_1} \left(\dfrac{1}{\delta_2}\right)^{1/\delta_2} \left(\dfrac{1}{\delta_3}\right)^{1/\delta_3} \\ \textit{sujeto a } \begin{cases} \delta_1, \delta_2, \delta_3 > 0 \\ \delta_1 + \delta_2 + \delta_3 = 1 \\ -\delta_1 + \delta_2 + \delta_3 = 0 \\ -\delta_1 + \delta_2 = 0 \end{cases} \end{array}\right\}$$

Al resolver este sistema se consigue solo la solución $\delta_1 = \delta_2 = \frac{1}{2}$ *y* $\delta_3 = 0$ *que no es factible para* (PG) *(pues* $\delta_3 = 0$*). Así, por el paso 2, el programa primario no tiene solución.*

La Programación geométrica restringida no será tratada en este trabajo. No obstante, se comenta a continuación las ideas fundamentales.

Dados $n, m \in \mathbb{N}$ dos números naturales, un **programa geométrico estándar** viene descrito por

$$\left.\begin{array}{c}\text{minimizar } g(\boldsymbol{t}) \text{ sujeto a} \\ h_1(\boldsymbol{t}) \leq 1, ..., h_m(\boldsymbol{t}) \leq 1\end{array}\right\} (PG)$$

donde tanto g como h_i $(i = 1, ..., m)$ son posinomios en n variables positivas $t_1, ..., t_n \in \mathbb{R}^+$.

Por un lado, al igual que ocurría en la Programación geométrica sin restricciones, resulta de utilidad la desigualdad aritmético-geométrica; en especial, una extensión suya que se prueba a partir de la propia Desigualdad aritmético-geométrica y cuya demostración puede verse en [40]. Por otro lado, a priori, el teorema 17 no parece ser aplicable en este caso puesto que los posinomios no tienen porque ser funciones convexas (por ejemplo, $g(t) = t^{1/2}$ no lo es). Sin embargo, todo posinomio puede transformarse en una función convexa a través el cambio de variable

$$t_j = e^{x_j} \qquad \forall j = 1, ..., n$$

Mas precisamente, mediante este cambio de variable se logra convertir el posinomio

$$g(\boldsymbol{t}) = \sum_{i=1}^{n} \alpha_i \prod_{j=1}^{m} (t_j)^{\beta_{ij}} \qquad \forall \boldsymbol{t} \in (\mathbb{R}^+)^n$$

en la siguiente función de $\boldsymbol{x} = (x_1, ..., x_n) \in \mathbb{R}^n$,

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i e^{\sum_{j=1}^{m} \beta_{ij} x_j}$$

que es convexa en $\mathbb{R}^n$ por la proposición 3. Estas observaciones permiten transformar el programa (PG) en el siguiente programa convexo

$$\left.\begin{array}{c}\text{minimizar } f(\boldsymbol{x}) \text{ sujeto a} \\ h_1^*(\boldsymbol{x}) - 1 \leq 0, ..., h_m^*(\boldsymbol{x}) - 1 \leq 0\end{array}\right\} (PG)^*$$

y ambos programas son equivalentes, en el sentido en que $\boldsymbol{t^0}$ es una solución de (PG) si, y sólo si, $\boldsymbol{x^0}$ es una solución de $(PG)^*$. En esta situación, si es aplicable el teorema 17, del cual se desprende un resultado análogo al teorema 19 y que puede consultarse en el capítulo 5 de [40].

3.4.3. Programación cuadrática

Los problemas de optimización no lineal en los que se optimiza una función cuadrática sujeta a restricciones lineales y de no negatividad, se presentan en una amplia variedad de aplicaciones incluyendo el análisis de la regresión en estadística o modelos económicos para la compra-venta óptima, entre otros. Se han desarrollado varios algoritmos para aprovecharse de la linealidad de las restricciones y del carácter cuadrático de la función

objetivo en tales programas. A continuación, se presenta el programa cuadrático estandar (el cual se resuelve con unos de estos algoritmos) y un caso particular que lograremos resolver por completo.

Dados $n, m \in \mathbb{N}$ dos números naturales, el **programa cuadrático estándar** se formula como sigue

$$\left.\begin{array}{c}\text{minimizar } f(x) = \kappa + \langle \boldsymbol{a}, \boldsymbol{x}\rangle + \frac{1}{2}\langle \boldsymbol{x}, Q\boldsymbol{x}\rangle \\ \text{sujeto a } A\boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0}\end{array}\right\} \quad (P)$$

donde $Q \in \mathfrak{M}_n(\mathbb{R})$ es una matriz definida positiva, $\boldsymbol{a} \in \mathbb{R}^n$ y $\boldsymbol{b} \in \mathbb{R}^m$ dos vectores, $\kappa \in \mathbb{R}$ un número real y $A \in \mathfrak{M}_{m\times n}(\mathbb{R})$ una matriz de rango máximo.

Este programa puede resolverse usando una variación del método simplex conocida como el *algoritmo de Wolfe* que puede verse en el capítulo 7 de [40]. Además, dado $n \in \mathbb{N}$ un número natural, puede verse (capítulo 3 de [40]) que el problema de minimizar la función cuadrática $f(\boldsymbol{x}) = \kappa + \langle \boldsymbol{a}, \boldsymbol{x}\rangle + \frac{1}{2}\langle \boldsymbol{x}, Q\boldsymbol{x}\rangle$ en todo $\mathbb{R}^n$ es equivalente a encontrar la única solución del sistema de ecuaciones lineal $A\boldsymbol{x} = -\boldsymbol{b}$, siendo $Q \in \mathfrak{M}_n(\mathbb{R})$ una matriz definida positiva, $\boldsymbol{a} \in \mathbb{R}^n$ y $\boldsymbol{b} \in \mathbb{R}^n$ dos vectores, $\kappa \in \mathbb{R}$ un número real y $A \in \mathfrak{M}_n(\mathbb{R})$ una matriz regular.

Ahora, vamos a considerar una clase de programas cuadráticos para los que la dualidad resulta fructífera. Dados $n \in \mathbb{N}$ un número natural, $Q \in \mathfrak{M}_n(\mathbb{R})$ una matriz simétrica y definida positiva, $\boldsymbol{a} \in \mathbb{R}^n$ un vector no nulo y $b \in \mathbb{R}$ un número real, considérese el programa cuadrático elemental

$$\left.\begin{array}{c}\text{minimizar } f(\boldsymbol{x}) = \frac{1}{2}\langle \boldsymbol{x}, Q\boldsymbol{x}\rangle \\ \text{sujeto a } \langle \boldsymbol{a}, \boldsymbol{x}\rangle \leq b\end{array}\right\} \quad (PQ)$$

La función $h : \mathbb{R}^n \to \mathbb{R}$ definida como $h(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x}\rangle - b$ es convexa y la función objetivo f es estrictamente convexa (por ser Q definida positiva), por tanto, (PQ) es un programa convexo con unicidad de solución. De hecho, (PQ) es superconsistente ya que por ser $\boldsymbol{a} \neq \boldsymbol{0}$, permite encontrar $\boldsymbol{x}$ con $\langle \boldsymbol{a}, \boldsymbol{x}\rangle < b$. Evidentemente, si $b \geq 0$, entonces $\boldsymbol{x_0} = \boldsymbol{0}$ es un punto factible para (PQ) y es la única solución del programa. Por esta razón, en lo que sigue se retringe la atención al caso $b < 0$. El lagrangiano $\mathcal{L} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ de (PQ) está dado por

$$\mathcal{L}(\boldsymbol{x}, \lambda) = \frac{1}{2}\langle \boldsymbol{x}, Q\boldsymbol{x}\rangle + \lambda(\langle \boldsymbol{a}, \boldsymbol{x}\rangle - b) \qquad \forall \boldsymbol{x} \in \mathbb{R}^n,\ \forall \lambda \in \mathbb{R}$$

Como el lagrangiano $\mathcal{L}$ es una función estrictamente convexa en $\boldsymbol{x}$ (para cada $\lambda \geq 0$ fijo), $\mathcal{L}$ tiene un único mínimo $\boldsymbol{x_0}$ que queda determinado por

$$\nabla_x \mathcal{L}(\boldsymbol{x}, \lambda) = Q\boldsymbol{x_0} + \boldsymbol{\lambda a} = \boldsymbol{0}$$

de donde

$$\boldsymbol{x_0} = -\lambda Q^{-1}\boldsymbol{a}$$

Además, como

$$\begin{aligned}
\mathcal{L}(\boldsymbol{x_0}, \lambda) &= \mathcal{L}(-\lambda Q^{-1}\boldsymbol{a}, \lambda) = \\
&= \frac{1}{2}\langle -\lambda Q^{-1}\boldsymbol{a}, Q(-\lambda Q^{-1}\boldsymbol{a})\rangle + \lambda(\langle \boldsymbol{a}, -\lambda Q^{-1}\boldsymbol{a}\rangle - b) = \\
&= \frac{1}{2}\langle -\lambda Q^{-1}\boldsymbol{a}, -\lambda\boldsymbol{a})\rangle - \langle -\lambda\boldsymbol{a}, -\lambda Q^{-1}\boldsymbol{a}\rangle - \lambda b = \\
&= -\frac{\lambda^2}{2}\langle Q^{-1}\boldsymbol{a}, \boldsymbol{a}\rangle - \lambda b
\end{aligned}$$

es claro que

$$\mathfrak{S} = \sup_{\lambda \geq 0}\left\{\inf_{\boldsymbol{x}\in\mathbb{R}^n} \mathcal{L}(\boldsymbol{x}, \lambda)\right\} = \max_{\lambda \geq 0}\left[-\frac{\lambda^2}{2}\langle Q^{-1}\boldsymbol{a}, \boldsymbol{a}\rangle - \lambda b\right]$$

Para maximizar la función $\varphi : \mathbb{R}_0^+ \to \mathbb{R}$ dada por

$$\varphi(\lambda) = -\frac{\lambda^2}{2}\langle Q^{-1}\boldsymbol{a}, \boldsymbol{a}\rangle - \lambda b \qquad \forall \lambda \in \mathbb{R}_0^+$$

basta escribir

$$\varphi'(\lambda) = -\lambda\langle Q^{-1}\boldsymbol{a}, \boldsymbol{a}\rangle - b$$
$$\varphi''(\lambda) = -\langle Q^{-1}\boldsymbol{a}, \boldsymbol{a}\rangle$$

Nótese que $\lambda_0 = -\dfrac{b}{\langle Q^{-1}\boldsymbol{a}, \boldsymbol{a}\rangle}$ es el único punto crítico de φ y es un máximo estricto de φ ya que $\varphi''(\lambda_0) < 0$. Así pues, λ_0 es una solución del programa dual. La solución $\boldsymbol{x_0}$ del programa primal es el mínimo de $\mathcal{L}$ en $\mathbb{R}^n$ y está dado por

$$\boldsymbol{x_0} = -\lambda_0 Q^{-1}\boldsymbol{a} = \frac{bQ^{-1}\boldsymbol{a}}{\langle \boldsymbol{a}, Q^{-1}\boldsymbol{a}\rangle}$$

Esto completa la solución de la clase de programas cuadráticos considerados.

Consúltese [35] para más información acerca de otras clases programación, como por ejemplo la *Programación separable* o la *Programación fraccionaria.*

Bibliografía

[1] Beltrami, E.J.: *A constructive proof of the Kuhn-Tucker multiplier rule*, Journal of Mathematical Analysis and Applications, Vol. 26, 1969, págs. 297–306.

[2] Kjeldsen, T. H.: *A Contextualized Historical analysis of the Kuhn-Tucker Theorem in Nonlinear Programming: The Impact of World War II*, Historia Mathematica, Vol. 27, 2000, págs. 331-361.

[3] Kjeldsen, T. H.: *A Contextualized Mathematico-Historical Analysis of Nonlinear Programming: Development and Multiple Discovery*, IMFUFA, text 372, Roskilde University, 1999.

[4] Rico Domínguez, E. M.: *Algunas desigualdades del Análisis matemático.* Trabajo de fin de grado supervisado por Facenda Aguirre J. A. (Dpto. Análisis Matemático) por la Universidad de Sevilla, 2017.

[5] Apostol, T. M.: *Análisis Matemático*, Reverté, 1960.

[6] Birbil, S. I., Frenk, J. G. B., Still, J. G.: *An elementary proof of the Fritz-John and Karush–Kuhn–Tucker conditions in nonlinear programming.* The European Journal of Operational Research, 2006 (article in press). Disponible en: *http://research.sabanciuniv.edu/177/1/3011800000548.pdf*

[7] Brinkhuis, J. y Protasov, V.: *A new proof of the Lagrange multiplier rule*, Operations Research Letters, Vol. 44, 2016, págs. 400-402.

[8] Acosta, M. D.; Aparicio, C.; Moreno, A. y Villena, A. R.: *Apuntes de Análisis Matemático I*, Universidad de Granada. Disponible en: *http://www.ugr.es/∼dpto_am/miembros/aparicio/apuntes/apuntes-an-mat-i-1-11-06.pdf*

[9] Brezhneva, O.; Tret'yakov, A. A. y Stephen, E. W.: *A short elementary proof of the Lagrange multiplier theorem*, Optimization Letters, Vol. 6, 2012, págs. 1597–1601.

[10] Ritoré, Manuel Mª: *Geometría de conjuntos convexos*, UGR. Disponible en *http://www.ugr.es/∼jperez/docencia/GeomConvexos/geometria_convexos-v2.pdf*

[11] Fletcher, R.: *Constrained Optimization*, Wiley, New York, 1981.

[12] van Tiel, J.: *Convex Analysis*, John Wiley and Sons, 1984.

[13] Neumann, J. V.: *Discussion of a maximum problem*, John Von Neumann Collected Works (Oxford) Vol. 6, págs. 89-95, 1947.

[14] Berkovitz L. D., Fleming H. W.: *Edward James McShane: a biographical memoire*. The National Academy Press, Vol. 80, 2001. Disponible en: *http://www.nasonline.org/publications/biographical-memoirs/memoir-pdfs/mcshane-e-j.pdf*

[15] Sominskii, I. S.: *El método de la inducción matemática*, Editorial Limusa, 1972.

[16] John, F.: *Extremum Problems with Inequalities as Subsidiary Conditions*, Presented to R. Courant on his 60th Birthday January (New York), Vol. 8, págs. 187-204, 1948.

[17] Paya Albert R.: *Funciones convexas*. Apuntes para la asignatura Cálculo II del primer curso del grado en Matemáticas, UGR. Disponible en *http://www.ugr.es/∼rpaya/documentos/CalculoII/2012-13/Convexas.pdf*

[18] Manzano Prego, J. M.: *Geometría de convexos*, UGR. Disponible en *http://www.ugr.es/∼jmmanzano/material/TallerGeoTop-Tema2.pdf*

[19] Franksen, O. I.: *Irreversibility by inequality constraints. I. On Fourier's inequality*, Systems Analysis Modelling, Vol. 2, págs. 137-149, 1985.

[20] O'Connor, J. J. y Robertson, E. F.: *Joseph-Louis Lagrange*, 1999. Disponible en: *http://www-history.mcs.st-andrews.ac.uk/Biographies/Lagrange.html*.

[21] Paya, R.: *Lección 8: Acotación y compacidad*. Apuntes de Análisis Matemático I del grado de Matemáticas (UGR) del curso 2014/5. Disponible en: *http://www.ugr.es/ rpaya/documentos/CalculoI/2012-13/Bolzano.pdf*

[22] Dantzig, G. B.: *Linear Programming and Extensions*, RAND Corporation, 1963.

[23] Sydsaeter, K. y Hammond, P.: *Matemáticas para el Análisis económico*, Pretince Hall, 1996.

[24] Lagrange, J-L.: *Méchanique analitique*, Reprint ed., París, 1965.

[25] Karush, W. *Minima of Functions of Several Variables with Inequalities as Side Conditions*, Departamento de Matemáticas (Universidad de Chicago), 1939.

[26] Pourciau, B.H.:*Modern multiplier rules*, The American Mathematical Monthly. Vol. 87, 1980, págs. 433–452.

[27] Kuhn, H. W. y Tucker, A. W.: *Nonlinear programming*, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (Berkeley), J. Neyman ed., págs. 481-492, 1950.

[28] Mangasarian, O. L.: *Nonlinear Programming*, SIAM, 1994.

[29] Kuhn, H. W.: *Nonlinear programming: a historical view*, SIAM-AMS Proceedings, Vol. 9, págs. 1-26, 1976.

[30] Bazaraa, M. S; Sherali, H. D. y Shetty, C. M.: *Nonlinear Programming, Theory and Algorithms*, 2ª ed. New York: Wiley, 1993.

[31] Bliss, G. A.: *Normality and abnormality in the calculus of variations*, Transactions of the American Mathematical Society, Vol. 43, págs. 365-376, 1938.

[32] Prékopa, A.: *On the development of optimation theory*, American Mathematical Monthly, Vol. 87, págs. 527-542, 1980.

[33] Clarke, Frank H.: *Optimization and Non-Smooth Analysis*, Wiley-Interscience, New York, 1983.

[34] Martin, Suárez, M.: *Orígenes del Cálculo diferencial e integral*. Apuntes de la asignatura Historia de las Matemáticas en el grado en Matemáticas, UGR. Disponible en *www.ugr.es/ ∼ mmartins/material/historia_matematica_origenes_calculo.pdf*

[35] Cantu Cuellar, R.: *Programación no lineal*, Universidad autónoma de Nueva León, 1996. Disponible en *http://eprints.uanl.mx/393/1/1080072405.PDF*.

[36] Cozzens, S.: *Social Control and Multiple Discovery in Science: The Opiate Receptor Case*, Satate University of New York Press, 1989.

[37] Kuhn, H. W.: *Subgrup theorems for groups Presented by generators and Relations*, Annals of Mathematics, Vol. 56, págs. 22-46, 1952.

[38] Paya, R.: *Tema 13: Propiedades de las funciones continuas*. Apuntes de Cálculo del grado de Matemáticas (UGR) del curso 2012/3. Disponible en: *http://www.ugr.es/ rpaya/documentos/CalculoI/2012-13/Bolzano.pdf*

[39] Sears, F. W. y Salinger, G. L.: *Termodinámica, teoría cinética y termondinámica estadiística*, Reverté, 1978.

[40] Peressini, E. A. L.; Sullivan, F. E. y Uhl, J. J.: *The Mathematics of Nonlinear Programming*, Springer-Verlag, 1988.

[41] McShane, J.: *The Lagrange multiplier rule*, The American Mathematical Monthly, Vol. 80, 1973, págs. 922-925.

[42] Farkas, G. J.: *Theorie der einfachen Ungleichungen*, Journal für die reine und angewadte Mathematik, Vol. 124, págs. 1-27, 1901.

[43] Hancock, H.: *Theory of maxima and minima*, Boston: Ginn, 1917.

[44] *Various proofs of the Cauchy-Schwarz inequality*, Octogon mathematical magazine, vol. 17, págs. 221-229, 2009.