UNIVERSIDAD NACIONAL DEL ALTIPLANO

FACULTAD DE INGENIERÍA CIVIL Y ARQUITECTURA

ESCUELA PROFESIONAL DE CIENCIAS FÍSICO MATEMÁTICAS

MÉTODO LAGRANGIANO AUMENTADO PARA LA SOLUCIÓN DE PROBLEMAS DE OPTIMIZACIÓN NO LINEAL

TESIS

PRESENTADA POR:

Bach. YENI ELIZABETH LAURACIO QUISPE

PARA OPTAR EL TÍTULO PROFESIONAL DE:

LICENCIADA EN CIENCIAS FÍSICO MATEMÁTICAS: MATEMÁTICAS

PUNO - PERÚ

2024

Reporte de similitud

NOMBRE DEL TRABAJO

MÉTODO LAGRANGIANO AUMENTADO PARA LA SOLUCIÓN DE PROBLEMAS DE OPTIMIZACIÓN NO LINEAL.pdf

AUTOR

YENI ELIZABETH LAURACIO QUISPE

RECUENTO DE PALABRAS

14401 Words

RECUENTO DE CARACTERES

62576 Characters

RECUENTO DE PÁGINAS

77 Pages

TAMAÑO DEL ARCHIVO

6.1MB

FECHA DE ENTREGA

Jul 25, 2024 9:47 AM GMT-5

FECHA DEL INFORME

Jul 25, 2024 9:48 AM GMT-5

13% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 12% Base de datos de Internet
- 4% Base de datos de publicaciones
- Base de datos de Crossref
- Base de datos de contenido publicado de Crossref
- 8% Base de datos de trabajos entregados

Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 12 palabras)
- Bloques de texto excluidos manualmente

UNIVERSIDAD NACIONAL DEL ALTIPLANO
Coordinador de Investigación de la E.P.F.M.
COORDINADOR
PUNO

Adelaida Otazu Conza
LIC. CIENCIAS FÍSICO - MATEMÁTICAS
Dra. MATEMÁTICA APLICADA
Subdirectora de Investigación de C.S Físico matemáticas

Mirsa Dolores Cruz Cuentas
DNI: 29615597

Resumen

DEDICATORIA

A mis padres Manuel y Fermina por ser los pilares fundamentales en mi vida, por haberme traído a este mundo, inculcarme buenos valores y darme la mejor educación. A mis abuelos maternos Zenobio y Emeteria por su cariño y enseñanzas de vida.

Yeni Elizabeth Lauracio Quispe

AGRADECIMIENTOS

Agradezco a Dios quien ha sido mi guía y fortaleza.

Mi especial gratitud al Dr. Ariel Rogelio Velazco Cárdenas por su asesoramiento y apoyo incondicional.

Asimismo agradezco a todos los docentes de la Escuela Profesional de Ciencias Físico Matemáticas de la Universidad Nacional del Altiplano que compartieron sus saberes y conocimientos a lo largo de mi formación profesional.

ÍNDICE GENERAL

TEMA: Método Lagrangiano Aumentado

ÁREA: Optimización No Lineal

LÍNEA DE INVESTIGACIÓN: Matemática Aplicada

FECHA DE SUSTENTACIÓN: 30 de julio del 2024

ÍNDICE DE FIGURAS

ÍNDICE DE ACRÓNIMOS

Símbolos

- $\|\cdot\|$: Norma cualquiera.
- $\|\cdot\|_2$: Norma Euclideana.
- $\{z_k\}$: Sucesión.
- C: Conjunto convexo.
- S: Conjunto cerrado, no vacío y convexo.
- $o(t)$: Orden.
- $\mathscr{A}(x)$: Conjunto activo.
- $\mathscr{C}, K$: Cono.
- $T_\Omega(x^*)$: Cono tangente.
- $\mathscr{F}(x)$: Cono viable linealizado.
- λ_i^*: Multiplicador de Lagrange.
- $Q(x;\mu)$: Función de Penalización Cuadrática.
- $\mathscr{L}_A(x,\lambda)$: Función Lagrangiana estándar del problema (4.1).
- $\mathscr{L}_A(x,\lambda;\mu)$: Función Lagrangiana Aumentada del problema (4.45).
- $\widetilde{\mathscr{L}_A}((x,t),\lambda;\mu)$: Función Lagrangiana Aumentada del problema (4.66).
- $\widehat{\mathscr{L}_A}((\tilde{x},t),\lambda;\mu)$: Función Lagrangiana Aumentada del problema (4.68).

Siglas y Abreviaciones

- PNL: Problemas de Optimización no Lineal.
- LICQ: Condiciones de Calificación de Independencia Lineal.
- KKT: Condiciones de Karush Kuhn Tucker.

RESUMEN

La presente investigación tiene por finalidad mostrar que el Método Lagrangiano Aumentado permite la solución del problema general de optimización no lineal. Para tal efecto se describe las condiciones de Karush Kuhn Tucker para puntos estacionarios y se estudia las Condiciones de Optimalidad de primer y segundo orden para soluciones locales. Se analiza el Método de Penalización cuadrática como base para el Método Lagrangiano Aumentado demostrándose que los puntos estacionarios del Lagrangiano Aumentado son puntos estacionarios del problema general de optimización no lineal y que, bajo ciertas condiciones el algoritmo generado por el Método Lagrangiano Aumentado converge a puntos estacionarios cuando es aplicado a un problema general de optimización no lineal, permitiendo así encontrar su solución. El algoritmo del Lagrangiano Aumentado genera una sucesión de iteraciones $\{x_k\}$ donde x_k es la solución aproximada de un subproblema que implica la función Lagrangiana Aumentada.

Palabras Clave: Condiciones de Karush Kuhn Tucker, Condiciones de Optimalidad, Lagrangiano Aumentado, Multiplicador de Lagrange, Optimización No Lineal, Puntos Estacionarios.

ABSTRACT

The purpose of this research is to show that the Augmented Lagrangian Method allows the solution of the general nonlinear optimization problem. For this purpose, the Karush Kuhn Tucker conditions for stationary points are described and the first and second order optimality conditions for local solutions are studied. The Quadratic Penalty Method is analyzed as a basis for the Augmented Lagrangian Method showing that the stationary points of the Augmented Lagrangian are stationary points of the general nonlinear optimization problem and that, under certain conditions the algorithm generated by the Augmented Lagrangian Method converges to stationary points when applied to a general nonlinear optimization problem, allowing thus to find its solution. The Augmented Lagrangian algorithm generates a sequence of iterations $\{x_k\}$ where x_k is the approximate solution of a subproblem involving the Augmented Lagrangian function.

Key Words: Karush Kuhn Tucker Conditions, Optimality Conditions, Augmented Lagrangian, Lagrangian Multiplier, Nonlinear Optimization, Stationary Points.

CAPÍTULO I

INTRODUCCIÓN

1.1. PLANTEAMIENTO DEL PROBLEMA

La investigación sobre problemas de optimización lineal cubren una amplia gama de aplicaciones, pero en general en la vida real las personas suelen encontrarse con otro tipo de problemas de optimización que son de tipo no lineales; la solución de estos problemas es de complejo desarrollo más aún si presentan restricciones de igualdad y desigualdad; por lo que se busca métodos o algoritmos de optimización eficientes para hallar la solución de estos problemas.

1.2. FORMULACIÓN DEL PROBLEMA

¿Será posible mostrar que, bajo ciertas condiciones, el algoritmo generado por el Método de Lagrangiano Aumentado converge a puntos estacionarios del problema de optimización no lineal con restricciones?

1.3. HIPÓTESIS DE LA INVESTIGACIÓN

1.3.1. Hipótesis General

El Método de Lagrangiano Aumentado permite la solución de problemas de optimización no lineal con restricciones de igualdad y desigualdad.

1.3.2. Hipótesis Específicas

1. Las condiciones de Karush Kuhn Tucker (KKT) se aplican para encontrar puntos estacionarios de problemas de optimización no lineal.

2. Las condiciones de optimalidad de primer y segundo orden garantizan que las condiciones de Karush Kuhn Tucker (KKT) generan soluciones locales.

1.4. JUSTIFICACIÓN DEL ESTUDIO

La investigación en el área de optimización no lineal es de suma importancia, puesto que es un principio básico del análisis de problemas complejos de decisión; además hay un amplio y creciente planteamiento de estos problemas en la teoría de control, en problemas de bioequivalencia, en la estadística aplicada, en la mecánica de fluidos, economía, estudio de poblaciones, etc. (Beck, 2023).

El motivo del presente trabajo es hacer estudios cualitativos de carácter explicativo y descriptivo de la solución del problema de optimización no lineal con restriciones de igualdad y desigualdad, las cuales tienen un alto grado de complejidad de desarrollo a comparación de los lineales.

Como consecuencia se ve necesario recurrir al Método Lagrangiano Aumentado que es asociado a los multiplicadores de Lagrange, este método desencadena un algoritmo que sirve para dar solución a problemas de optimización con restricciones.

Por lo expuesto, para el desarrollo de la presente investigación, es preciso brindar aspectos teóricos partiendo de conceptos, teoremas y propiedades primordiales.

1.5. OBJETIVOS DE LA INVESTIGACIÓN

1.5.1. Objetivo General

Mostrar que, bajo ciertas condiciones, el algoritmo generado por el Método de Lagrangiano Aumentado converge a puntos estacionarios del problema de optimización no lineal con restricciones, que son soluciones locales.

1.5.2. Objetivos Específicos

1. Describir las condiciones de Karush Kuhn Tucker (KKT) para problemas de optimización no lineal.

2. Estudiar condiciones de optimalidad de primer y segundo orden para garantizar la validez de las condiciones de Karush Kuhn Tucker (KKT).

3. Estudiar métodos de penalización para hallar la solución de problemas de optimización no lineal.

4. Mostrar que el Método Lagrangiano Aumentado genera puntos estacionarios del problema de optimización no lineal con restricciones.

CAPÍTULO II

REVISIÓN DE LITERATURA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

En los problemas de optimización no lineal se pueden presentar algunos donde sea necesario maximizar una función no lineal, sin embargo como $\max f(x) = -\min(-f(x))$ se considera solo problemas de minimización en esta investigación; así pues presentamos algunos estudios de investigadores que de una u otra manera tienen que ver con la investigación planteada.

Mehta (2020) halla la solución del problema de optimización con restricciones no lineales de objetivo único mediante un algoritmo iterativo para las múltiples topologías y muestra así el proceso de convergencia global, donde los resultados numéricos se obtienen por la herramienta de solución Octave.

Yi-Chih, Yung-Cheng, & Peng-Sheng, (2015) proponen un método artificial y mejorado que consta de dos fases tomando de referencia el método inmunoevolutivo para la resolución de problemas de optimización restringida no lineal que presenta variables reales, variables enteras y variables discretas.

Wang, Zhang, Wang, & Rong (2020) diseñan un algoritmo de primer orden para encontrar una solución óptima inexacta de problemas de optimización no lineal restringidos utilizando un software computacional económico, apoyándose de una actualización de parámetros de penalización durante la resolución de los subproblemas generados.

Bertsekas (1999) proporciona una información completa y accesible de algoritmos para resolver problemas de optimización no lineal, centrándose en un análisis matemático riguroso. Hace una introducción de los métodos modernos y sus aplicaciones.

Nocedal & Wright (1999) brindan una descripción completa de las técnicas más poderosas y de vanguardia para resolver problemas de optimización; muestra las diferentes aplicaciones que tienen desarrollando de manera comprensible diferentes algoritmos y señala el camino a las investigaciones futuras sobre la mejora y ampliación de los algoritmos en softwares de optimización.

Rojas (2018) estudia las condiciones de cualificación y optimización del problema de relaciones análogas del equilibrio de Nash generalizado mediante el algoritmo de Lagrangiano Aumentado, utiliza las condiciones de términos para probar la convergencia global y a su vez calcular un punto de Karush-Kuhn Tucker (KKT).

Croceri & Sottosanto (2011) muestran la resolución de problemas de cuadrados mínimos no lineales con restricciones de igualdad y desigualdad mediante un algoritmo apoyado en la minimización secuencial del Lagrangiano Aumentado, se apoyan con el método de gradiente proyectado y el esquema de tipo Armijo; a su vez presentan resultados numéricos previos mediante una inserción en el software Matlab.

Cui, Ding, Li, & Zhao (2021) aplican el Método de Lagrangiano Aumentado para la resolución de problemas de optimización de matrices convexas de gran escala, analizan dos tipos de condiciones suficientes para asegurar las condiciones de crecimiento cuadrático de una clase de problemas de optimización restringidas y regularizadas por funciones espectrales no suaves. Finalmente hacen una introducción de la resolución de problemas de optimización de la matriz convexa central.

Kanzow, Steck, & Wachsmuth (2018) proponen una variación del Método de Lagrangiano Aumentado para resolucion de problemas de optimización restringidas en espacios de Banach, además presentan soluciones numéricas para realzar la factibilidad del método.

Soledad (2018) extiende el Método Lagrangiano Aumentado para la resolución de problemas de optimización multiobjetivo restringidos mediante el uso de escalarizaciones y el uso de la función Lagrangiana Aumentada vectorial; se demuestra la convergencia a un punto estacionario del problema multiobjetivo.

Andreani, Secchin, Ramos, Ribeiro & Velazco (2021) afirman que el algoritmo de Lagrangiano Aumentado es un método exitoso para resolver problemas de optimización con restricciones, donde el análisis de convergencia es mejor con el empleo de noción de condiciones de optimización secuencial.

Vázquez (2021) presenta la resolución de problemas de optimización con restricciones mediante métodos de penalización y el Método de Lagrangiano Aumentado; afirma que este último es un método de perfeccionamiento respecto a los métodos anteriores ya que tiene una mayor velocidad de convergencia, de esta forma hace una introducción a los algoritmos de resolución de problemas de optimización con restricciones. Por consiguiente es considerado como base para la presente investigación.

2.2. MARCO TEÓRICO

En esta sección veremos algunos conceptos y notaciones propias relacionadas con el problema:

$$\min_{x} f(x) \text{ sujeto a } \begin{cases} c_i(x) = 0, \ i \in \varepsilon \\ \\ c_i(x) \geq 0, \ i \in I \end{cases}$$

donde $f, c_i : \mathbb{R}^n \to \mathbb{R}$ y f es conocida como función objetivo; además I y ε conjuntos finitos de índices.

Sea n un número natural. El espacio vectorial a considerar es $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ veces}}$ donde los elementos de dicho espacio son vectores columna. Como en $\mathbb{R}^n$ todas las normas son equivalentes, $\|\cdot\|$ denota una norma cualquiera en $\mathbb{R}^n$ y $\|\cdot\|_2$ denota la norma euclideana. El producto interior en $\mathbb{R}^n$ se denota por $x^T y$, para todo $x, y \in \mathbb{R}^n$ y la sucesión será denotada por $\{z_k\}$.

2.2.1. Conceptos Topológicos

En esta subsección nos ocuparemos de introducir los conceptos básicos topológicos a manera de fijar notaciones.

Definición 2.2.1. *Una* ***sucesión*** *en* $\mathbb{R}^n$ *es una aplicación* $k \in \mathbb{N} \rightarrow z_k \in \mathbb{R}^n$*, definida en el conjunto* $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ *denotada por* $\{z_k\}$.

Decimos que el punto $\bar{x} \in \mathbb{R}^n$ es el **límite de la sucesión** $\{z_k\}$ cuando para todo $\varepsilon > 0$ dado, es posible obtener $\bar{k} \in \mathbb{N}$ tal que

$$k \geq \bar{k} \Rightarrow \|z_k - \bar{x}\| < \varepsilon$$

En este caso, también decimos que la sucesión $\{z_k\}$ **converge** para $\bar{x}$ y representamos esto por $z_k \rightarrow \bar{x}$ o $\lim\limits_{k\rightarrow\infty} z_k = \bar{x}$ o $\lim z_k = 0$ si no hubiese confusión.

En particular una sucesión $\{z_k\}$ **satisface la propiedad** P **para** k **suficientemente grande**, si existe N tal que P se satisface para todo $k > N$. (Alves & Wegner, 2014)

Por ejemplo si $z_k \rightarrow 5$ entonces significa que $\|z_k - 5\| < 1$ o que $\|z_k\| < 10$.

Existen resultados importantes respecto a la sucesión que nos servirán para el desarrollo del presente trabajo, los cuales resaltamos a continuación:

- Si una sucesión $\{z_k\}$ converge para un límite $\bar{x}$, entonces toda subsucesión $\{z_{k_i}\}_{i\in\mathbb{N}}$ también converge para $\bar{x}$.

- Por el Teorema de Bolzano Weierstrass toda sucesión limitada en $\mathbb{R}^n$ posee una subsucesión convergente.

- Sean las sucesiones $\{z_k\} \subset \mathbb{R}^n$ y $\{\lambda_k\} \subset \mathbb{R} - \{0\}$, con $\lambda_k \to 0$. Se dice que z_k es el orden de λ_k, es decir $z_k = o(\lambda_k)$ cuando $\dfrac{z_k}{\lambda_k} \to 0$; y asociado a este concepto se tiene que:
    - $o(t_k) + o(t_k) = o(t_k)$.
    - $b^T o(t_k) = o(t_k)$.
    - $o\big(o(t_k)\big) = o(t_k)$.
    - $o(\|o(t_k)\|) = o(t_k)$.

(Lages, 2004)

Definición 2.2.2. *Dado el punto $a \in \mathbb{R}^n$ y el número real $r > 0$, la **bola abierta** de centro a y radio r es el conjunto $B(a;r)$ de los puntos $x \in \mathbb{R}^n$ cuya distancia al punto a es menor que r. En símbolos:*

$$B(a;r) = \{x \in \mathbb{R}^n; \|x - a\| < r\}$$

*Análogamente, la **bola cerrada** de centro a y radio r es el conjunto $B[a;r]$ así definido:*

$$B[a;r] = \{x \in \mathbb{R}^n; \|x - a\| \leq r\}$$

Se dice que el conjunto $X \subset \mathbb{R}^n$ es limitado cuando está contenido en alguna bola $B[a;r]$. Como $B[a;r] \subset B[0;k]$, donde $k = r + |a|$, decir que X es limitado equivale a decir $k > 0$ tal que $\|x\| \leq k$ para todo $x \in X$.

Una función $f : X \to \mathbb{R}^n$ se dice limitada cuando su imagen $f(X) \subset \mathbb{R}^n$ es un conjunto limitado, es decir, cuando existe $c > 0$ tal que $\|f(x)\| \leq c$ para todo $x \in X$. (Lages, 2004)

Definición 2.2.3. *Sea $a \in X \subset \mathbb{R}^n$, se dice que el punto a es interior al conjunto X cuando, para algún $r > 0$ se tiene $B(a,r) \subset X$. Esto significa que todos los puntos suficientemente* *próximos de a también pertenecen a X. El conjunto interior X de los* ***puntos interiores*** *a X se llama* ***interior*** *del conjunto X, denotado por $\operatorname{int}X$. Evidentemente, $\operatorname{int}X \subset X$. Cuando $a \in \operatorname{int}X$, se dice que X es una vecindad de a.Un conjunto $A \subset \mathbb{R}^n$ se llama* ***abierto*** *cuando todos sus puntos son interiores, es decir cuando $A = \operatorname{int}A$. (Lages, 2004).*

Definición 2.2.4. *Un punto a es* ***adherente*** *al conjunto $X \subset \mathbb{R}^n$ cuando existe una sucesión de puntos x_k en X tal que $\lim x_k = a$. (Lages, 2004).*

Definición 2.2.5. *Se llama* ***clausura*** *del conjunto $X \subset \mathbb{R}^n$ al conjunto $\overline{X}$ formado por todos los puntos adherentes a X. Por tanto $a \in \overline{X}$ si y solo si $a = \lim z_k$, $z_k \in X$. Se dice que $a \in \overline{X}$ es lo mismo afirmar que a es adherente a X.Un conjunto $F \in \mathbb{R}^n$ se llama* ***cerrado*** *cuando $\overline{F} = F$ es decir, cuando el límite de toda sucesión convergente de puntos de F es un punto de F.*

Entre los resultados topológicos importantes que vamos a utilizar, podemos destacar:

- Un conjunto $F \subset \mathbb{R}^n$ es cerrado si y solamente si, su complemento $\mathbb{R}^n - F$ es abierto. Equivalentemente $A \subset \mathbb{R}^n$ es abierto si y solamente si $\mathbb{R}^n - A$ es cerrado.

- La clausura de cualquier conjunto $X \subset \mathbb{R}^n$ es cerrado. Es decir para todo $X \subset \mathbb{R}^n$ se tiene $\overline{\overline{X}} = \overline{X}$.

- Sea $F \subset \mathbb{R}^n$ un conjunto cerrado. Dado cualquier $a \in \mathbb{R}^n$ existe por lo menos un $x_0 \in F$ tal que $\|x_0 - a\| \leq \|x - a\|$ para todo $x \in X$.

- Un conjunto $X \subset \mathbb{R}^n$ se llama **compacto** cuando es limitado y cerrado. Las siguientes afirmaciones sobre el conjunto $K \subset \mathbb{R}^n$ son equivalentes:

  - K es compacto.

  - Toda sucesión de puntos $z_k \in K$ posee una subsucesión que converge para un punto de K.

- Sea $K \subset \mathbb{R}^n$ compacto y $F \subset \mathbb{R}^n$ cerrado. Existen $x_0 \in K$ y $y_0 \in F$ tal que $\|x_0 - y_0\| \leq \|x - y\|$ para cualquier $x \in K$ y $y \in F$.

(Lages, 2004)

Definición 2.2.6. *Se dice que $a \in \mathbb{R}^n$ es el **punto de acumulación** del conjunto $X \subset \mathbb{R}^n$ cuando toda bola de centro a contiene algún punto de X diferente de a.*

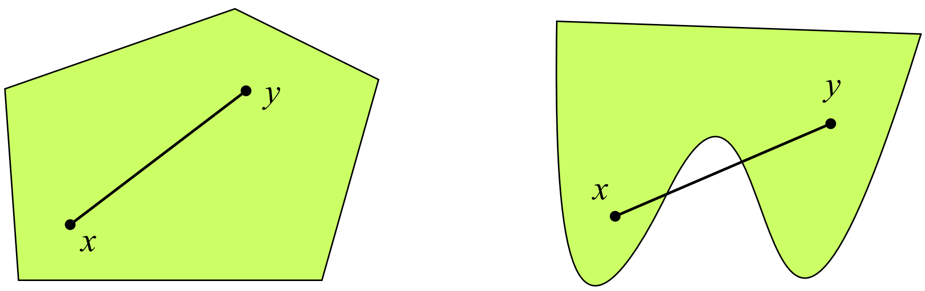
Los resultados topológicos resaltantes que utilizaremos con respecto al punto de acumulación son:

- La equivalencia de las siguientes afirmaciones: a es un punto de acumulación de X, a es el límite de una sucesión de puntos $z_k \in X - \{a\}$ y toda bola de centro a contiene una infinidad de puntos de X.

- Todo **subconjunto infinito limitado** $X \subset \mathbb{R}^n$ admite por lo menos un punto de acumulación.

(Alves & Wegner, 2014)

Definición 2.2.7 (Conjunto Convexo)**.** *Un conjunto $C \subset \mathbb{R}^n$ es llamado **convexo** cuando $\forall x, y \in C$, el segmento $[x, y] = \{(1-t)x + ty \mid t \in [0, 1]\}$ está enteramente contenido en C. (Alves & Wegner, 2014)*



Convexo   No Convexo

Figura 2.2.1: Conjunto Convexo y No Convexo.

Lema 2.2.1. *Sean $u,v \in \mathbb{R}^n$ con $u \neq v$. Si $\|u\|_2 = \|v\|_2 = r$, entonces $\|(1-t)u+tv\|_2 < r$, para todo $t \in (0,1)$. (Alves & Wegner, 2014)*

Lema 2.2.2. *Sea $S \subset \mathbb{R}^n$ un conjunto cerrado no vacío. Dado $z \in \mathbb{R}^n$, existe $\bar{z} \in S$ tal que*

$$\|z-\bar{z}\| \leq \|z-x\|$$

para todo $x \in S$. (Alves & Wegner, 2014)

Lema 2.2.3. *Sea $S \subset \mathbb{R}^n$ un conjunto no vacío, convexo y cerrado. Dado $s \in \mathbb{R}^n$, existe un único $\bar{s} \in S$ tal que*

$$\|s-\bar{s}\|_2 \leq \|s-x\|_2$$

para todo $x \in S$. (Alves & Wegner, 2014)

2.2.2. Funciones Diferenciables

Sea $f : U \to \mathbb{R}$ una función definida en el abierto $U \subset \mathbb{R}^n$. Para cada $i = 1,\ldots,n$, la i-ésima derivada parcial de f en el punto $a = (a_1,\ldots,a_n) \in U$ es el número

$$\frac{\partial f}{\partial x_i} = \lim_{t\to 0} \frac{f(a+te_i)-f(a)}{t} = \lim_{t\to 0} \frac{f(a_1,\ldots,a_i+t,\ldots,a_n)-f(a)}{t}$$

caso este límite exista.

Sea $f : U \to \mathbb{R}$ una función que posee las n derivadas parciales en todos los puntos del abierto $U \subset \mathbb{R}^n$. Luego se definen n funciones.

$$\frac{\partial f}{\partial x_i},\ldots,\frac{\partial f}{\partial x_n} : U \to \mathbb{R}, \text{ donde } \frac{\partial f}{\partial x_i} : x \mapsto \frac{\partial f}{\partial x_i}(x)$$

Si estas funciones son continuas en U, diremos que f es una función de **clase $\mathbf{C^1}$** y escribiremos $f \in C^1$.

Una función $f : U \to \mathbb{R}^n$, definida en el abierto $U \subset \mathbb{R}^n$, se dice de clase C^1 cuando cada una de sus funciones coordenadas $f_1,\ldots,f_n : U \to \mathbb{R}$ es de clase C^1.

Una función $f:U\to\mathbb{R}$, definida en el abierto $U\subset\mathbb{R}^n$, se dice **diferenciable** en el punto $a\in U$ cuando cumple las siguientes condiciones:

1. Existen las derivadas parciales $\dfrac{\partial f}{\partial x_1}(a),\ldots,\dfrac{\partial f}{\partial x_n}(a)$.

2. Para todo $v=(\alpha_1,\ldots,\alpha_n)$ tal que $a+v\in U$, se tiene

$$f(a+v)-f(a)=\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\cdot\alpha_i+r(v)\,,\ \text{donde}\ \lim_{\|v\|\to 0}\frac{r(v)}{\|v\|}=0$$

Una función $f:U\to\mathbb{R}^n$, definido en el abierto $U\subset\mathbb{R}^n$, se dice diferenciable en el punto $a\in U$ cuando cada una de sus funciones coordenadas $f_1,\ldots,f_n:U\to\mathbb{R}$ es diferenciable en ese punto.

(Lages, 2004).

Teorema 2.2.1. *Toda función $f:U\to\mathbb{R}$ de clase C^1 es diferenciable.*

Sea $f:U\to\mathbb{R}$ una función que posee las derivadas parciales $\dfrac{\partial f}{\partial x_1}(x),\ldots,\dfrac{\partial f}{\partial x_n}(x)$ en todo punto x del abierto $U\subset\mathbb{R}^n$. La j-ésima derivada parcial de la función $\dfrac{\partial f}{\partial x_i}:U\to\mathbb{R}$ en el punto $x\in U$ será indicada por

$$\frac{\partial^2 f}{\partial x_j\partial x_i}(x)=\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x)\,,\quad i,j=1,\ldots,n$$

Si estas derivadas parciales de segundo orden existieran en cada punto $x\in U$ tendremos n^2 funciones $\dfrac{\partial^2 f}{\partial x_j\partial x_i}:U\to\mathbb{R}$. Cuando tales funciones fueran continuas diremos que f es de **clase $\mathbf{C^2}$** y escribiremos $f\in C^2$. (Lages, 2004)

Considerando $f:\mathbb{R}^n\to\mathbb{R}$ una función de clase C^2 entonces su **gradiente** y **hessiana** se definen como:

$$\nabla f=\begin{pmatrix}\dfrac{\partial f}{\partial x_1}\\ \vdots\\ \vdots\\ \dfrac{\partial f}{\partial x_n}\end{pmatrix};\qquad \nabla^2 f=\begin{pmatrix}\dfrac{\partial^2 f}{\partial x_1\partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_1\partial x_n}\\ \vdots & \ddots & \vdots\\ \dfrac{\partial^2 f}{\partial x_n\partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n\partial x_n}\end{pmatrix}$$

(Alves & Wegner, 2014)

Ejemplo 2.2.1. *Sea $f:\mathbb{R}^2 \to \mathbb{R}$ dada por*

$$f(x) = 2x_1^2 - x_1x_2 + x_2^2 - 2x_1 + \frac{2}{3}x_2 + e^{x_1+x_2}$$

- *La gradiente de f es:*

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 4x_1 - x_2 - 2 + e^{x_1+x_2} \\ -x_1 + 2x_2 + \dfrac{2}{3} + e^{x_1+x_2} \end{pmatrix}$$

- *La hessiana de f es:*

$$\nabla^2 f(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2 \partial x_2} \end{pmatrix} = \begin{pmatrix} 4 + e^{x_1+x_2} & -1 + e^{x_1+x_2} \\ -1 + e^{x_1+x_2} & 2 + e^{x_1+x_2} \end{pmatrix}$$

Teorema 2.2.2 (Teorema de la Función Implícita)**.** *Dada una función $f:U \to \mathbb{R}$, de clase C^k $(k \geq 1)$ en el abierto $U \subset \mathbb{R}^{n+1}$, sea $(x_0, y_0) \in U$ tal que $f(x_0,y_0) = c$ y $\dfrac{\partial f}{\partial y}(x_0,y_0) \neq 0$. Existen una bola $B = B(x_0;\delta)$ y un intervalo $J = (y_0 - \varepsilon, y_0 + \varepsilon)$ con las siguientes propiedades:*

a) $B \times \bar{J} \subset U$ *y* $\dfrac{\partial f}{\partial y}(x,y) \neq 0$ *para todo* $(x,y) \in B \times \bar{J}$.

b) *Para todo $x \in B$ existe un único $y = \xi(x) \in J$ tal que $f(x,y) = f(x,\xi(x)) = c$.*

Una función $\xi : B \to J$, así definida es de clase C^k y sus derivadas parciales en cada punto $x \in B$ son dadas por

$$\frac{\partial \xi}{\partial x_i} = \frac{-\frac{\partial f}{\partial x_i}(x, \xi \partial x)}{\frac{\partial f}{\partial y}(x, \xi(x))}$$

(Lages, 2004)

Teorema 2.2.3 (Teorema de Taylor)**.** *Sea $f:\mathbb{R}^n \to \mathbb{R}$ una función diferenciable, si $\bar{x} \in \mathbb{R}^n$, entonces*

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + r(x)$$

con $\lim_{x \to \bar{x}} \frac{r(x)}{\|x-\bar{x}\|} = 0$. *(Alves & Wegner, 2014)*

Corolario 2.2.1. *Sea* $f : \mathbb{R}^n \to \mathbb{R}$ *es una función diferenciable, si* $\bar{x} \in \mathbb{R}^n$ *y* $x = \bar{x} + td$ *entonces*

$$f(\bar{x}+td) = f(\bar{x}) + t\nabla f(\bar{x})^T d + o(t). \tag{2.1}$$

(Alves & Wegner, 2014)

Teorema 2.2.4 (Teorema de Taylor de Segundo Orden)**.** *Si* $f : \mathbb{R}^n \to \mathbb{R}$ *es una función dos veces diferenciable y* $\bar{x} \in \mathbb{R}$*, entonces*

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x-\bar{x}) + \frac{1}{2}(x-\bar{x})^T \nabla^2 f(\bar{x})(x-\bar{x}) + r(x)$$

con $\lim_{x \to \bar{x}} \frac{r(x)}{\|x-\bar{x}\|^2} = 0$. *(Alves & Wegner, 2014)*

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. MATERIALES

En el presente trabajo de investigación, los materiales que se utilizaron fueron:

- Libros.
- Revistas indexadas.
- Artículos.
- Papers online.
- Laptop.
- Internet.
- Memoria USB.
- Papel bond.
- Lapiceros.

3.2. MÉTODOS

El presente trabajo "Método Lagrangiano Aumentado para la solución de problemas de opimización no lineal", debido a Charmaz (2008) se desarrolló utilizando el método descriptivo dentro del contexto de la teoría fundamentada, también se utiliza la hermenéutica que permite comprender, interpretar y evaluar la hipótesis y objetivos de la presente investigación.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. OPTIMIZACIÓN CON RESTRICCIONES

Consideremos el problema general de optimización no lineal con restricciones.

$$\min_{x} f(x) \ \text{sujeto a} \ \begin{cases} c_i(x) = 0, & i \in \varepsilon \\ c_i(x) \geq 0, & i \in I \end{cases} \tag{4.1}$$

donde:

- $f, c_i : \mathbb{R}^n \to \mathbb{R}$ y f es conocida como función objetivo.
- I y ε conjuntos finitos de índices, donde:
  - c_i, $i \in \varepsilon$ son las restricciones de igualdad.
  - c_i, $i \in I$ son las restricciones de desigualdad.

Definiéndose el **conjunto viable**

$$\Omega = \{x \mid c_i(x) = 0, i \in \varepsilon; c_i(x) \geq 0, i \in I\} \tag{4.2}$$

como el conjunto de puntos x que satisfacen las restricciones.

Para este caso de problemas con restricciones nos limitaremos a los puntos viables en una vecindad de x^*.

Un vector x^* será una **solución local del problema** (4.1) si $x^* \in \Omega$ y existe una vecindad $\mathscr{N}$ de x^* tal que $f(x) \geq f(x^*)$ para $x \in \mathscr{N} \cap \Omega$.

El **conjunto activo** $\mathscr{A}(x)$ para cualquier punto viable x, se compone del conjunto ε junto con los índices de las restricciones de desigualdad para las cuales $c_i(x) = 0$; es decir

$$\mathscr{A}(x) = \varepsilon \cup \{i \in I \mid c_i(x) = 0\} \tag{4.3}$$

Observación: En un punto viable x,

- La restricción de desigualdad $i \in I$ es activa si $c_i(x) = 0$.
- La restricción de desigualdad $i \in I$ es inactiva si $c_i(x) > 0$

4.1.1. Cono Tangente y Cono Viable Linealizado

En esta subsección definiremos los conceptos de conos y la Condición de Calificación de Independencia Lineal (LICQ) los cuales serán de utilidad para la demostración de las condiciones de Karush Kuhn Tucker (KKT).

Definición 4.1.1 (Cono)**.** *Un subconjunto no vacío $\mathscr{C} \subset \mathbb{R}^n$ es un* **cono**, *si para todo $t \geq 0$ y $d \in \mathscr{C}$ se tiene que $td \in \mathscr{C}$.*

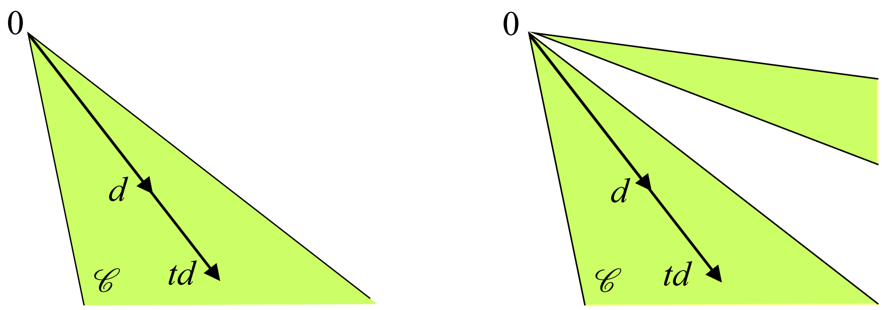


Figura 4.1.1: Conos.

Ejemplo 4.1.1 (Cono convexo y cerrado)**.** *Dados los vectores $v_1, v_2, v_3, \cdots, v_m \in \mathbb{R}^n - \{0\}$, el conjunto*

$$\mathscr{C} = \left\{ \sum_{i=1}^{m} y_i v_i \mid y_i \geq 0, i = 1, \cdots, m \right\}$$

demostrar que es un cono convexo y cerrado.

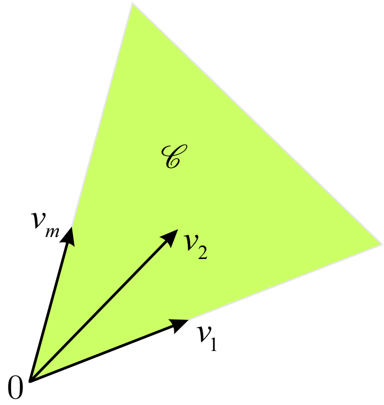


Figura 4.1.2: Figura del Ejemplo 4.1.1

Demostración.

Considerando la matriz $B = \begin{pmatrix} v_1 & v_2 & \cdots & v_m \end{pmatrix} \in \mathbb{R}^{n\times m}$, tenemos

$$\mathscr{C} = \{By \mid y \in \mathbb{R}^m, y \geq 0\}$$

Para mostrar que $\mathscr{C}$ es cono, tomemos $d = By \in \mathscr{C}$ y $t \geq 0$. Asimismo, $td = B(ty) \in \mathscr{C}$, pues $ty \geq 0$. La convexidad se obtiene de la relación $(1-t)By + tBw = B\big((1-t)y + tw\big)$. Ahora probaremos que $\mathscr{C}$ es cerrado, por inducción en m.

i) Para $m = 1$. Sea $(d^k) \subset \mathscr{C}$, tal que $d^k \to d$. Tenemos $d^k = y^k v_1$, con $y^k \geq 0$. Asimismo,

$$||v_1||^2 y^k = v_1^T \to v_1^T d$$

lo que implica que $y^k \to y$, donde $y = \dfrac{v_1^T d}{||v_1||^2} \geq 0$, pues $y^k \geq 0$. Por tanto, $d^k = y^k v_1 \to y v_1$ y asimismo, $d = y v_1 \in \mathscr{C}$.

ii) Suponiendo que el ejemplo sea válido para $m-1$. Vamos a demostrar que vale para m. Considerando primero el caso que $rango(B) = m$. Sea $(d^k) \subset \mathscr{C}$, tal que $d^k \to d$. Entonces, $d^k = By^k$, con $y^k \geq 0$. De este modo

$$B^T B y^k = B^T d^k \to B^T d$$

así $y^k \to y$, con $y = (B^T B)^{-1} B^T d$. Como $y^k \geq 0$, tenemos $y \geq 0$. Por lo tanto, $d^k = By^k \to By$ y asi, $d = By \in \mathscr{C}$. Supongamos que $rango(B) < m$; las columnas de B son linealmente dependientes, lo que implica que existe $\gamma \in \mathbb{R}^m$ tal que

$$B\gamma = 0 \tag{4.4}$$

y $\gamma > 0$ para algún $i = 1, \cdots, m$. Considerando, para $j = 1, \cdots, m$, la matriz

$$B_j = \begin{pmatrix} v_1 & \cdots & v_{j-1} & v_{j+1} & \cdots & v_m \end{pmatrix} \in \mathbb{R}^{n \times (m-1)}$$

obtenida suprimiendo la j-ésima columna de B. Por inducción, tenemos que el conjunto

$$C_j = \left\{ B_j z \mid z \in \mathbb{R}^{M-1},\, z \geq 0 \right\}$$

es cerrado para todo $j = 1, \cdots, m$. Por lo tanto, la unión $\cup_j^m C_j$ es un conjunto cerrado. Para concluir la demostración, vamos a mostrar que $C = \cup_j^m C_j$. Sea $d \in C$. Entonces $d = By$, para algún $y \geq 0$. Considerando

$$\bar{t} = \text{máx} \left\{ -\frac{y_i}{\gamma_i} \mid \gamma^i > 0 \right\}$$

donde γ es dado por (4.4). Asi, para todo i tal que $\gamma_i > 0$, se tiene $y_i + t\gamma_i \geq 0$. Además como $\bar{t} \leq 0$, en consecuencia también vale $y_i + \bar{t}\gamma_i \geq 0$ para cada i tal que $\gamma_i \leq 0$. Sea j tal que $\bar{t} = -\dfrac{y_j}{\gamma_j}$. Definiendo $\bar{y} = y + \bar{t}\gamma$, tenemos que $\bar{y} \geq 0$ y $\bar{y}_j = 0$. Por lo tanto de la ecuación (4.4), obtenemos

$$d = By = B(y + \bar{t}\gamma) = B\bar{y} \in C_j$$

Ya que la inclusión $\cup_j^m C_j \subset C$ es inmediata, completamos la prueba. □

Definición 4.1.2. *Sea un punto viable x, $\{z_k\}$ es una* **sucesión viable** *que converge a x si $z_k \in \Omega$ para todo k suficientemente grande y $z_k \to x$.*

Definición 4.1.3. *Un vector d es* **tangente** *a Ω en un punto x^* si existe una sucesión viable $\{z_k\}$ en Ω que converge a x^* y una sucesión de escalares positivos $\{t_k\}$ con $t_k \to 0$ de modo que*

$$\lim_{k\to\infty} \frac{z_k - x^*}{t_k} = d \tag{4.5}$$

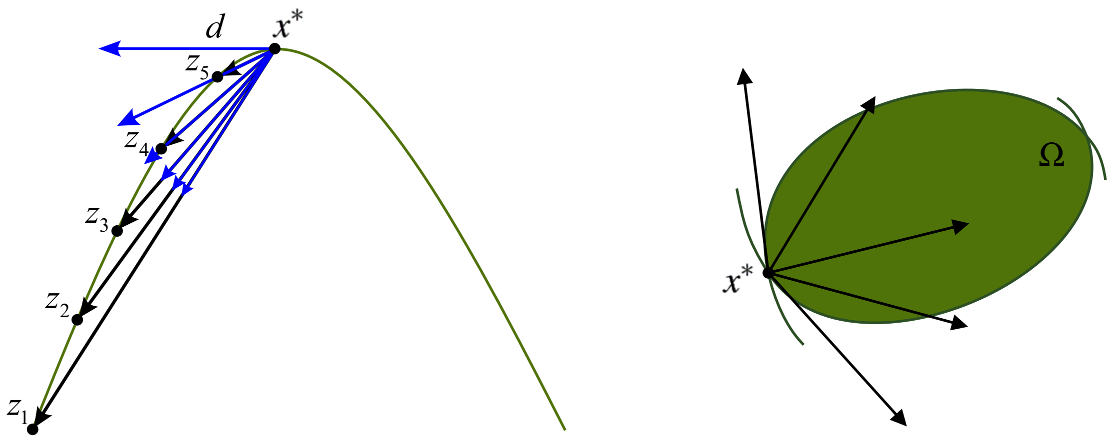


Figura 4.1.3: Direcciones Tangentes.

el conjunto de los vectores tangentes a Ω en x^* se denomina **cono tangente** *y es denotado por $T_\Omega(x^*)$.*

Definición 4.1.4. *El* **cono viable linealizado** *se define como el conjunto*

$$\mathscr{F}(x^*) = \left\{ d \;\middle|\; \begin{array}{ll} d^T \nabla c_i(x^*) = 0, & \forall i \in \varepsilon \\ d^T \nabla c_i(x^*) \geq 0, & \forall i \in \mathscr{A}(x^*) \cap I \end{array} \right\} \tag{4.6}$$

<u>Observación:</u> Note que el cono tangente y el cono viable linealizado satisfacen la definición de cono.

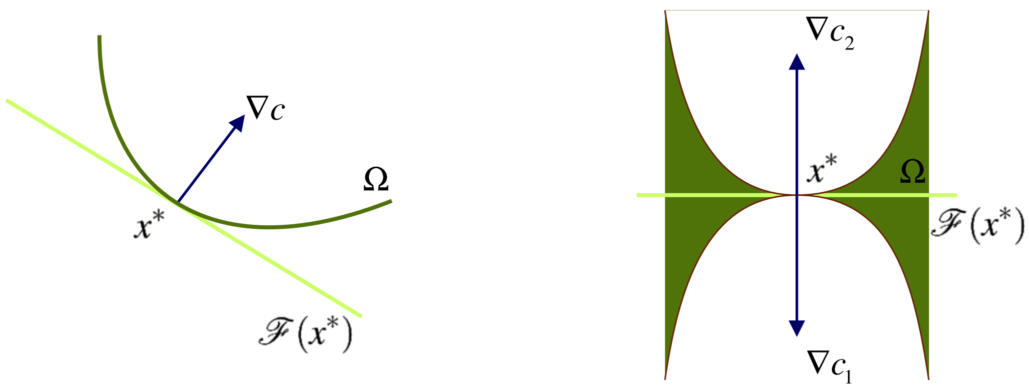


Figura 4.1.4: Conos Viables Linealizados.

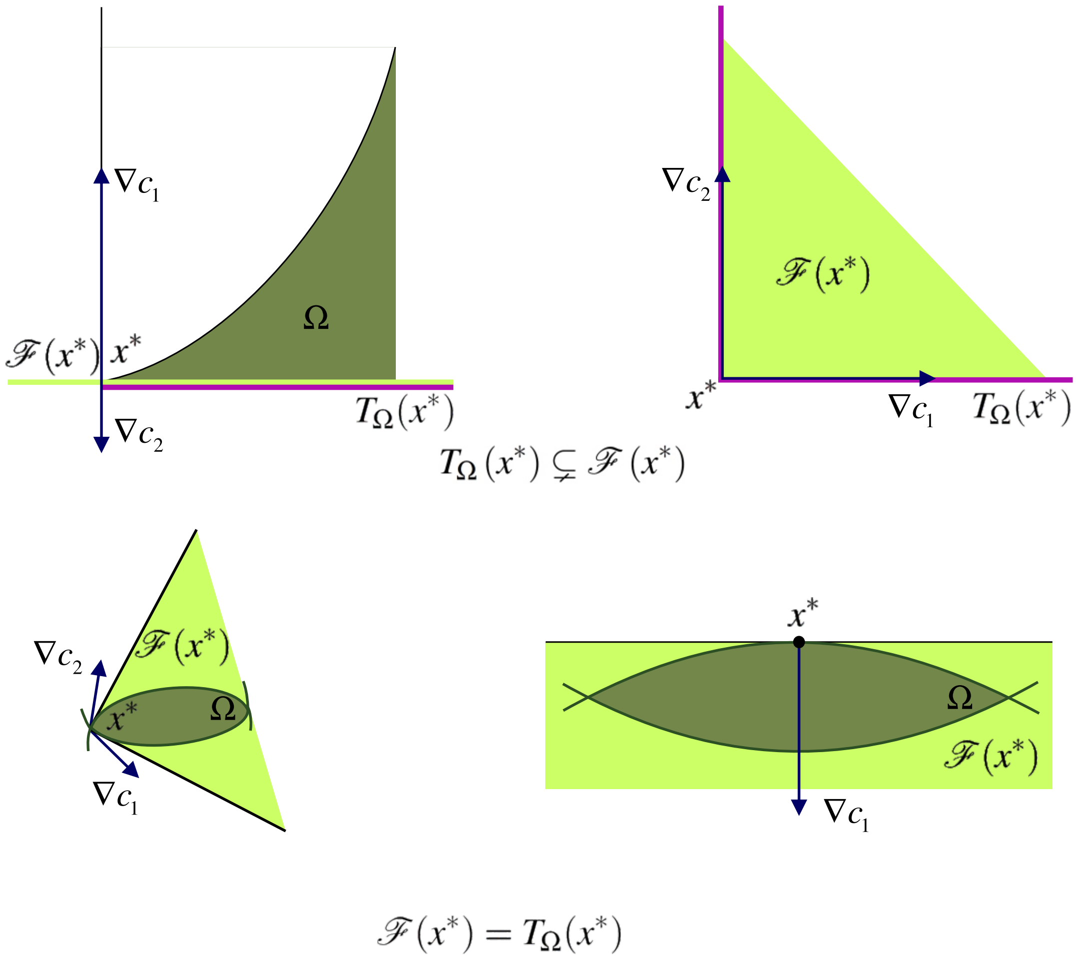


Figura 4.1.5: Conos Viables Linealizados-Conos Tangentes.

Definición 4.1.5 (LICQ)**.** *Dado el problema general (4.1), un punto $x \in \Omega$ satisface la* **Condiciones de Calificación de Independencia Lineal (LICQ)** *si el conjunto de gradientes de las restricciones activas $\{\nabla c_i(x^*), i \in \mathscr{A}(x^*)\}$ es linealmente independiente.*

4.1.2. Condiciones de Optimalidad de Primer Orden

Las condiciones de optimalidad de primer orden ayudarán a determinar la validez de las condiciones de Karush Kuhn Tucker (KKT), previamente para enunciar las condiciones necesarias de optimalidad definiremos la función Lagrangiana para el problema general (4.1) el cual se define de la siguiente forma.

$$\mathscr{L}(x,\lambda) = f(x) - \sum_{i\in\varepsilon\cup I} \lambda_i c_i(x) \tag{4.7}$$

Teorema 4.1.1 (Condiciones necesarias de primer orden)**.** *Sea x^* una solución local de (4.1) , donde f y c_i son funciones continuamente diferenciables, si x^* satisface LICQ, entonces existe un vector λ^* llamado multiplicador de lagrange, de componentes λ_i^* para $i \in \varepsilon \cup I$ tal que se cumplen las siguientes condiciones denominadas condiciones de Karush Kuhn Tucker (KKT):*

$$\nabla_x \mathscr{L}(x^*,\lambda^*) = 0 \tag{4.8a}$$

$$c_i(x^*) = 0, \forall i \in \varepsilon \tag{4.8b}$$

$$c_i(x^*) \geq 0, \forall i \in I \tag{4.8c}$$

$$\lambda_i^* \geq 0, \forall i \in I \tag{4.8d}$$

$$\lambda_i^* c_i(x^*) = 0, \forall i \in \varepsilon \cup I \quad \textit{(Condición de complementariedad)} \tag{4.8e}$$

como $\lambda_i^* = 0$ por la condición de complementariedad para todo $i \notin \mathscr{A}(x^*)$, entonces (4.8a), se reescribe como:

$$0 = \nabla_x \mathscr{L}(x^*,\lambda^*) = \nabla f(x^*) - \sum_{i\in\mathscr{A}(x^*)} \lambda_i^* \nabla c_i(x^*)$$

donde si se satisface LICQ entonces los $\nabla c_i(x^*)$ son linealmente independientes, luego λ^* es único. (Nocedal & Wright, 1999)

Para llegar a la demostración de este Teorema 4.1.1 necesitaremos de varios resultados previos.

Primeramente veremos una relación entre el cono tangente y el cono viable linealizado, para ello consideraremos a $A(x^*)$ como la matriz cuyas filas son los gradientes de las restricciones activas en el punto viable x^*, es decir:

$$A(x^*)^T = \left[\nabla c_i(x^*)\right]_{i\in\mathscr{A}(x^*)} \tag{4.9}$$

donde el conjunto activo $\mathscr{A}(x^*)$ se define como en la ecuación (4.3).

Lema 4.1.1. *Sea x^* un punto viable del problema general (4.1). Entonces las dos afirmaciones siguientes son verdaderas.*

i) $T_\Omega(x^*) \subset \mathscr{F}(x^*)$.

ii) Si la condición LICQ se satisface en x^*, entonces $\mathscr{F}(x^*) = T_\Omega(x^*)$.

Demostración.

Consideremos las restricciones activas c_i en x^*, con $i = 1, 2, \cdots, m$. Para (i) :

Si $d \in T_\Omega(x^*)$, además $\{z_k\}$ y $\{t_k\}$ sean las sucesiones para las que se cumple las condiciones de vector tangente (4.5), así,

$$\lim_{k\to\infty} \frac{z_k - x^*}{t_k} = d, \quad t_k > 0 \ \text{ para todo } \ k$$

es decir $\displaystyle\lim_{k\to\infty} \frac{z_k - x^* - dt_k}{t_k} = 0$ lo que implica que $z_k - x^* - dt_k = o(t_k)$. Luego, se tiene que

$$z_k = x^* + t_k d + o(t_k) \tag{4.10}$$

Tomando $i \in \varepsilon$ y luego por el Corolario 2.2.1, se tiene que

$$
\begin{aligned}
0 &= \frac{1}{t_k} c_i(z_k) \\
&= \frac{1}{t_k}\left[(c_i(x^*) + t_k d) + t_k \nabla(c_i(x^*) + t_k d) o(t_k) + o(\|o(t_k)\|)\right] \\
&= \frac{1}{t_k}\left[(c_i(x^*) + t_k \nabla c_i(x^*)^T d + o(t_k)) + t_k \nabla(c_i(x^*) + t_k d) o(t_k) + o(\|o(t_k)\|)\right] \\
&= \frac{1}{t_k}\left[c_i(x^*) + t_k \nabla c_i(x^*)^T d + o(t_k)\right] \\
&= \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}.
\end{aligned}
$$

luego considerando el límite cuando $k \to \infty$, se tiene que

$$\nabla c_i(x^*)^T d = 0 \tag{4.11}$$

Para las restricciones de desigualdad activas $i \in \mathscr{A}(x^*) \cap I$, de manera similar se tiene que

$$
\begin{aligned}
0 &\leq \frac{1}{t_k} c_i(z_k) \\
&\leq \frac{1}{t_k}\left[c_i(x^*) + t_k \nabla c_i(x^*)^T d + o(t_k)\right] \\
&\leq \nabla c_i(x^*)^T d + \frac{o(t_k)}{t_k}.
\end{aligned}
$$

y considerando el límite cuando $k \to \infty$ se tiene que

$$\nabla c_i(x^*)^T d \geq 0 \tag{4.12}$$

Luego por (4.11) y (4.12) se tiene que $d \in \mathscr{F}(x^*)$ por lo que $T_\Omega(x^*) \subset \mathscr{F}(x^*)$.

Para (ii):

Como LICQ se satisface en x^*, entonces la matriz $A(x^*)$ de dimensiones $m \times n$, formada por los gradientes de las restricciones activas tiene un rango fila completo m. Luego sea Z una matriz donde sus columnas son una base para el espacio nulo de $A(x^*)$, es decir

$$Z \in \mathbb{R}^{n\times(n-m)},\ Z \text{ tiene rango columna completo},\ A(x^*)Z = 0 \tag{4.13}$$

Sea $d \in \mathscr{F}(x^*)$ y suponiendo que $\{t_k\}_{k=0}^{\infty}$ es una sucesión de escalares positivos de modo que $\lim\limits_{k\to\infty} t_k = 0$. Consideremos el sistema de ecuaciones paramétricas R por

$$R(z,t) = \begin{bmatrix} c(z) - tA(x^*)d \\ Z^T(z - x^* - td) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4.14}$$

Afirmamos que las soluciones $z = z_k$ de este sistema para $t = t_k > 0$ dan una sucesión viable que tiende a x^* y que cumplen la Definición 4.1.3.

Para $t = 0$, $z = x^*$, el jacobiana de R en este punto es

$$\nabla_z R(x^*, 0) = \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix} \tag{4.15}$$

que resulta ser no singular por construcción de Z. En consecuencia por el Teorema de la función implícita (2.2.2), el sistema (4.13) tiene una solución única z_k para todos los valores de t_k suficientemente pequeños. Asimismo, de (4.14) y de la Definición 4.1.4 se tiene que

$$i \in \varepsilon \Rightarrow c_i(z_k) = t_k \nabla c_i(x^*)^T d = 0 \tag{4.16}$$

$$i \in \mathscr{A}(x^*) \cap I \Rightarrow c_i(z_k) = t_k \nabla c_i(x^*)^T d \geq 0 \tag{4.17}$$

por lo que z_k es viable.

Como $R(z_k, t_k) = 0$ para todo k, entonces por el Teorema de Taylor 2.2.3 y considerando (4.9) se tiene que

$$0 = R(z_k, t_k) = \begin{bmatrix} c(z_k) - t_k A(x^*)d \\ Z^T(z_k - x^* - t_k d) \end{bmatrix}$$
$$= \begin{bmatrix} A(x^*)(z_k - x^*) + o(||z_k - x^*||) - t_k A(x^*)d \\ Z^T(z_k - x^* - t_k d) \end{bmatrix}$$
$$= \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix} (z_k - x^* - t_k d) + o(||z_k - x^*||)$$

Note que como

$$\frac{z_k - x^* - t_k d}{t_k} = o\left(\frac{||z_k - x^*||}{t_k}\right)$$

$$0 = \frac{1}{t_k}\begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix} (z_k - x^* - t_k d) + \frac{1}{t_k} o(||z_k - x^*||)$$

$$0 = \frac{1}{t_k}(z_k - x^* - t_k d) + \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix}^{-1} \frac{o(||z_k - x^*||)}{t_k} \tag{4.18}$$

$$0 = \frac{z_k - x^*}{t_k} - d + \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix}^{-1} \frac{o(||z_k - x^*||)}{t_k}$$

luego

$$\frac{z_k - x^*}{t_k} = d - \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix}^{-1} \frac{o(||z_k - x^*||)}{t_k} \tag{4.19}$$

por otro lado

$$\lim_{k\to\infty} \frac{o(||z_k - x^*||)}{||z_k - x^*||} = 0$$

entonces

$$\lim_{k\to\infty} \frac{\frac{o(||z_k - x^*||)}{t_k}}{\frac{||z_k - x^*||}{t_k}} = 0$$

de donde

$$\frac{o(||z_k - x^*||)}{t_k} = o\left(\frac{||z_k - x^*||}{t_k}\right)$$

así reemplazando en (4.19)

$$
\begin{aligned}
\frac{z_k - x^*}{t_k} &= d - \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix}^{-1} \frac{o(||z_k - x^*||)}{t_k} \\
&= d + o\left(\frac{||z_k - x^*||}{t_k}\right)
\end{aligned}
$$

luego tomando el límite cuando $k \to \infty$ se tiene que

$$
\lim \frac{z_k - x^*}{t_k} = d
$$

por lo tanto, $d \in T_\Omega(x^*)$ para un d arbitrario $\in \mathscr{F}(x^*)$, es decir $\mathscr{F}(x^*) \subseteq T_\Omega(x^*)$ y considerando (i) del Lema 4.1.1 se tiene que $\mathscr{F}(x^*) = T_\Omega(x^*)$. $\square$

Teorema 4.1.2. *Si x^* es una solución local de (4.1), entonces se tiene*

$$
\nabla f(x^*)^T d \geq 0, \ \text{para todo} \ d \in T_\Omega(x^*) \tag{4.20}
$$

Demostración.

Sea $d \in T_\Omega(x^*)$, $d \neq 0$. Entonces existe una sucesión $z_k \subset \Omega$ tal que $z_k \to x^*$ y $\frac{z_k - x^*}{\|z_k - x^*\|} \to \frac{d}{\|d\|}$. Por otro lado, tenemos

$$
0 \leq f(z_k) - f(x^*) = \nabla f(x^*)^T (z_k - x^*) + o(\|z_k - x^*\|),
$$

para todo k suficientemente grande. Dividiendo por $\|z_k - x^*\|$ y tomando el límite obtenemos $\nabla f(x^*)^T d \geq 0$, completando la prueba. $\square$

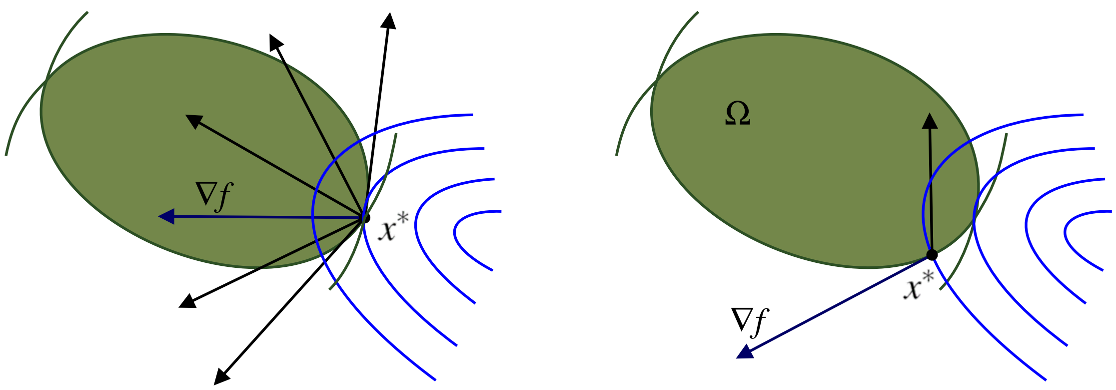


Figura 4.1.6: Regiones entre Direcciones Tangentes y el Gradiente de la Función Objetivo.

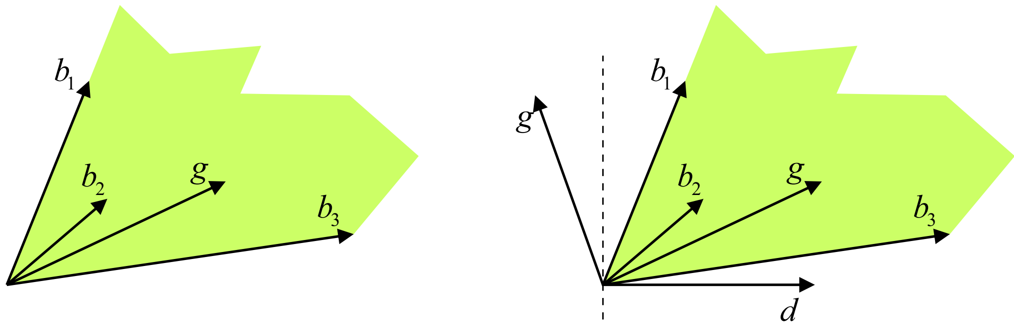
A continuación presentaremos un Lema que es indispensable para demostrar el Teorema 4.1.1. Este lema define un cono K como:

$$K = \{By + Cw \mid y \geq 0\} \tag{4.21}$$

donde B y C son matrices de dimensión $n \times m$ y $n \times p$, respectivamente, además y y w son vectores de dimensiones apropiadas. Sea un vector $g \in \mathbb{R}^n$, el Lema de Farkas afirma que bien $g \in K$, o bien existe un vector $d \in \mathbb{R}^n$ de modo que

$$g^T d < 0, \;\; B^T d \geq 0, \;\; C^T d = 0 \tag{4.22}$$

este lema asevera que una y sólo una de estas alternativas es cierta.



(a) O bien $g \in K$ (b) O bien existe un hiperplano de separación

Figura 4.1.7: Lema de Farkas.

Lema 4.1.2 (Lema de Farkas)**.** *Sea el cono K definido como en (4.21) y sea cualquier vector $g \in \mathbb{R}^n$, se tiene que $g \in K$ o que existe $d \in \mathbb{R}^n$ que satisface (4.22), pero no ambos. (Nocedal & Wright, 1999)*

Demostración.

Primero se demuestra que las dos opciones no pueden cumplirse a la misma vez. Si $g \in K$, existen vectores $y \geq 0$ y w de modo que $g = By + Cw$. Además si existe un d con la propiedad (4.22), por productos internos se tiene que

$$0 > d^T g = d^T By + d^T Cw = (B^T d)^T y + (C^T d)^T w \geq 0$$

donde de la última desigualdad se sigue el hecho de $C^T d = 0$, $B^T d \geq 0$, y $y \geq 0$. Por lo tanto, no se puede tener ambas alternativas a la vez.

Mostraremos ahora que una de las condiciones siempre se satisface. Primeramente construiremos d con la propiedad (4.22) en el caso que $g \notin K$ para ellos utilizaremos el hecho que K es cerrado.

Corolario 4.1.1. *Sea:*

$$C = \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_p \end{bmatrix}$$

siendo $c_i \in \mathbb{R}^{n\times 1}$ *(vectores columna) con* $i = 1, \cdots, p$

$$w = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_p \end{bmatrix}^T$$

Note que

$$Cw = \sum_{i=1}^{p} w_i c_i$$

$$Cw = \sum_{w_i > 0} w_i c_i + \sum_{w_i \leq 0} w_i c_i$$

$$Cw = \sum_{w_i > 0} w_i c_i + \sum_{w_i \leq 0} (-w_i)(-c_i)$$

$$\Rightarrow Cw = \sum_{i=1}^{p} v_i d_i \ \ ; v_i \geq 0$$

Luego sea:

$$B = \begin{bmatrix} b_1 & b_2 & b_3 & \cdots & b_m \end{bmatrix}$$

siendo $b_i \in \mathbb{R}^{n\times 1}$ *(vectores columna), con* $i = 1, \cdots, m$

$$y = \begin{bmatrix} y_1 & y_2 & y_3 & \dots & y_m \end{bmatrix}^T$$

$$By = \sum_{i=1}^{m} y_i b_i$$

Así,

$$K = By + Cw, y \geq 0$$

$$K = \sum y_i b_i + \sum v_i d_i,\ \ d_i b_i \geq 0$$

luego por el Ejemplo 4.1.1 tenemos que K es cerrado.

Como K es convexo y cerrado por los Lemas 2.2.2 y 2.2.3 existe un vector $\widehat{s}$ que está más próximo de g en el sentido de la norma euclidiana, es decir

$$||\widehat{s} - g||_2^2 = \min_{s\in K} ||s - g||_2^2 \tag{4.23}$$

Ya que $\widehat{s} \in K$, y dado que K es un cono se tiene que $\alpha\widehat{s} \in K$ para todos los escalares $\alpha \geq 0$. Además puesto que $||\alpha\widehat{s} - g||_2^2$ se minimiza con $\alpha = 1$, calculando se tiene que

$$\begin{aligned}
\frac{d}{d\alpha}||\alpha\widehat{s} - g||_2^2\Big|_{\alpha=1} &= \frac{d}{d\alpha}(\alpha\widehat{s} - g)^T(\alpha\widehat{s} - g) \\
&= \widehat{s}^T(\alpha\widehat{s} - g) + (\alpha\widehat{s} - g)^T\widehat{s} \\
&= \alpha\widehat{s}^T\widehat{s} - \widehat{s}^T g + \alpha\widehat{s}\widehat{s}^T - g^T\widehat{s} \\
&= (2\alpha\widehat{s}^T\widehat{s} - 2\widehat{s}^T g)|_{\alpha=1} = 0 \\
&= 2\widehat{s}^T\widehat{s} - 2\widehat{s}^T g = 0 \\
&= \widehat{s}^T(\widehat{s} - g) = 0
\end{aligned} \tag{4.24}$$

Ahora, sea s cualquier otro vector en K. Como K es convexo, por la propiedad (4.23) de minimización de $\widehat{s}$ se tiene que

$$||\widehat{s} + \theta(s - \widehat{s}) - g||_2^2 \geq ||\widehat{s} - g||_2^2 \text{ para todo } \theta \in (0,1]$$

donde

$$||\theta(s - \widehat{s}) + \widehat{s} - g||_2^2 = \theta^2||s - \widehat{s}||_2^2 + ||\widehat{s} - g||_2^2 + 2\theta(s - \widehat{s})^T(\widehat{s} - g) \geq ||\widehat{s} - g||_2^2$$

En consecuencia

$$2\theta(s-\widehat{s})^T(\widehat{s}-g)+\theta^2||s-\widehat{s}||_2^2 \geq 0$$

dividiendo esta última expresión por θ y tomando el límite a medida que $\theta \downarrow 0$, se tiene que $(s-\widehat{s})^T(\widehat{s}-g) \geq 0$. Por ende, a causa de (4.24)

$$s^T(\widehat{s}-g) \geq 0, \;\; \text{para todo} \;\; s \in K \tag{4.25}$$

Ahora afirmamos que el vector

$$d = \widehat{s}-g$$

cumple las condiciones (4.22). Se observa que $d \neq 0$ puesto que $g \notin K$. Así de (4.24) se tiene que

$$d^Tg = d^T(\widehat{s}-d) = (\widehat{s}-g)^T\widehat{s}-d^Td = -||d||_2^2 < 0,$$

luego d cumple la primera propiedad de (4.22).

Por otro lado de (4.25), se tiene que $d^Ts \geq 0$ para todo $s \in K$, entonces

$$d^T(By+Cw) \geq 0 \;\; \text{para todo} \;\; y \geq 0 \;\; \text{y todo} \;\; w.$$

Luego si $y = 0$ se tiene que $(C^Td)^Tw \geq 0$ para todo w, lo cual es válido solamente si $C^Td = 0$. Similarmente si $w = 0$, se tiene que $(B^Td)^Ty \geq 0$ para todo $y \geq 0$, lo cual es válido solamente si $B^Td \geq 0$. Por lo tanto, d también cumple la segunda y tercera propiedad en (4.22) lo que concluye la demostración. □

Corolario 4.1.2. *Aplicando el Lema de Farkas al cono N definido por las condiciones del problema general (4.1) se tiene*

$$N = \left\{ \sum_{i \in \mathscr{A}(x^*)} \lambda_i \nabla c_i(x^*), \, \lambda_i \geq 0 \;\; para \;\; i \in \mathscr{A}(x^*) \cap I \right\} \tag{4.26}$$

y considerando el vector $g \in N$

$$g = \nabla f(x^*)$$

existe d tal que

$$g = \nabla f(x^*) = \sum_{i \in \mathscr{A}(x^*)} \lambda_i^* \nabla c_i(x^*) = A(x^*)^T \lambda^*, \lambda_i^* \geq 0 \ para \ i \in \mathscr{A}(x^*) \cap I \tag{4.27}$$

con $A(x^*)$ definido como en (4.3). O bien existe una dirección d de modo que

$$d^T \nabla f(x^*) < 0$$

ya que

$$d^T \nabla c_i(x^*) \geq 0 \ para \ i \in \mathscr{A}(x^*) \cap I$$

$$d^T \nabla c_i(x^*) = i \in \varepsilon$$

lo que implica que $d \in \mathscr{F}(x^*)$ por la Definición 4.1.4.

Ahora para demostrar el Teorema 4.1.1, consideramos los Lemas 4.1.1 y 4.1.2 para generar las condiciones KKT definidas en el Teorema 4.1.1. Así pues sea $x^* \in \mathbb{R}^n$ una solución local en el que se cumplen la condición LICQ. Luego por el Teorema 4.1.2 se tiene

$$d^T \nabla f(x^*) \geq 0 \ \text{para todo} \ d \in \mathscr{F}(x^*)$$

Así por el Corolario 4.1.2 existe λ^* que satisface (4.27).

Finalmente afirmamos que el vector λ^* está definido por

$$\lambda_i^* = \begin{cases} \lambda_i, & i \in \mathscr{A}(x^*) \\ 0, & i \in I | \mathscr{A}(x^*) \end{cases} \tag{4.28}$$

satisface las condiciones de KKT (4.8) junto con la solución local x^*.

En efecto:

- La condición (4.8a) es derivada inmediatamente de (4.27) , la definición (4.7) de la función Lagrangiana y la definición de λ^* en (4.28).

- Ya que x^* es viable, se cumplen las condiciones (4.8b) y (4.8c).

- Se tiene de (4.27) que $\lambda_i^* \geq 0$ para $i \in \mathscr{A}(x^*) \cap I$, en tanto que de (4.28), $\lambda_i^* = 0$ para $i \in I | \mathscr{A}(x^*)$. Por ende, $\lambda_i^* \geq 0$ para $i \in I$, cumpliendo así (4.8d).

- Se tiene para $i \in \mathscr{A}(x^*) \cap I$ que $c_i(x^*) = 0$, en tanto que para $i \in I | \mathscr{A}(x^*)$, se tiene $\lambda_i^* = 0$. Por lo que $\lambda_i^* c_i(x^*) = 0$ para $i \in I$, cumpliendo así (4.8e).

Por lo tanto, la demostración está completa.

Definición 4.1.6 (Complementariedad Estricta)**.** *Dada una solución local x^* de (4.1) y el vector λ^* que satisface (4.8), la condición de complementariedad estricta se cumple si exactamente uno de λ_i^* y $c_i(x^*)$ es cero para el índice $i \in I$. En otras palabras tenemos $\lambda_i^* > 0$ para cada $i \in I \cap \mathscr{A}(x^*)$.*

4.1.3. Condiciones de Optimalidad de Segundo Orden

Por Taylor

$$f(x^* + w) = f(x^*) + \nabla f(x^*)^T w + o(w) \tag{4.29}$$

donde

$$\frac{f(x^* + w) - f(x^*) - \nabla f(x^*)^T w}{\|w\|} \longrightarrow 0$$

Luego si x^* es un minimizador local por el Teorema 4.1.2 $\nabla f(x^*)^T w \geq 0$ para todo $w \in \mathscr{F}(x^*)$. Así en (4.29) f crece si $\nabla f(x^*)^T w > 0$ y mantiene su valor si $\nabla f(x^*)^T w = 0$.

En consecuencia no podemos determinar si la función crece o decrece considerando solo información de primer orden, por lo que son primordiales las condiciones de segundo orden las cuales analizan los términos de la segunda derivada en las expansiones en serie de Taylor de f y c_i. Para ello se considera a f y c_i, $i \in \varepsilon \cup I$ dos veces continuamente diferenciables.

Dado el cono $\mathscr{F}(x^*)$ de la Definición 4.1.4 y un vector multiplicador de Lagrange λ^* que cumpla las condiciones KKT (4.8), se define el **cono crítico** $\mathscr{C}(x^*, \lambda^*)$ de la siguiente forma:

$$\mathscr{C}(x^*,\lambda^*) = \left\{w \in \mathscr{F}(x^*) | \nabla c_i(x^*)^T w = 0, \text{ para todo } i \in \mathscr{A}(x^*) \cap I \text{ con } \lambda_i^* > 0\right\}$$

De manera equivalente se tiene,

$$w \in \mathscr{C}(x^*,\lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0, & \text{para todo } i \in \varepsilon \\ \nabla c_i(x^*)^T w = 0, & \text{para todo } i \in \mathscr{A}(x^*) \cap I \text{ con } \lambda_i^* > 0 \\ \nabla c_i(x^*)^T w \geq 0, & \text{para todo } i \in \mathscr{A}(x^*) \cap I \text{ con } \lambda_i^* = 0 \end{cases} \tag{4.30}$$

El cono crítico contiene las direcciones w que tienden a pegarse a las restricciones de desigualdad activas, incluso cuando se hace una pequeña variación en el objetivo, aquellos índices $i \in I$ para los que las componentes λ_i^* del multiplicador de Lagrange sean positivas así como a las restricciones de igualdad. De la ecuación (4.30) y del caso que $\lambda_i^* = 0$ para todas las componentes inactivas $i \in I | \mathscr{A}(x^*)$, se tiene que

$$w \in \mathscr{C}(x^*,\lambda^*) \Rightarrow \lambda_i^* \nabla c_i(x^*)^T w = 0 \text{ para todo } i \in \varepsilon \cup I \tag{4.31}$$

por lo que, de la primera condición KKT (4.8a) y de (4.7) que define la función Lagrangiana, se tiene

$$w \in \mathscr{C}(x^*,\lambda^*) \Rightarrow w^T \nabla f(x^*) = \sum_{i \in \varepsilon \cup I} \lambda_i^* w^T \nabla c_i(x^*) = 0 \tag{4.32}$$

A continuación se define una condición necesaria que implica a las segundas derivadas. Si x^* es una solución local, entonces el hessiano de la Lagrangiana tiene una curvatura no negativa a lo largo de las direcciones en $\mathscr{C}(x^*, \lambda^*)$ llamadas direcciones críticas.

Teorema 4.1.3 (Condiciones necesarias de segundo orden)**.** *Sea x^* una solución local de (4.1) y que se cumple la condición LICQ. Si λ^* es un vector multiplicador de Lagrange para el que se cumplen las condiciones KKT de (4.8), entonces*

$$w^T \nabla^2_{xx} \mathscr{L}(x^*, \lambda^*) w \geq 0, \quad \textit{para todo } \; w \in \mathscr{C}(x^*, \lambda^*) \tag{4.33}$$

(Nocedal & Wright, 1999)

Demostración.

Utilizamos la idea de demostración del Lema 4.1.1 como en x^* se satisface la LICQ, entonces por el Lema 4.1.1 $\mathscr{C}(x^*, \lambda^*) \subset \mathscr{F}(x^*) = \mathscr{T}_\Omega(x^*)$ así si $w \in \mathscr{C}(x^*, \lambda^*)$ entonces existe sucesiones $\{z_k\}$ y $\{t_k\}$ con $z_k \longrightarrow x^*$ de modo que

$$\lim_{k \to \infty} \frac{z_k - x^*}{t_k} = w \tag{4.34}$$

donde (4.34) también se puede escribir como

$$z_k - x^* = t_k w + o(t_k) \tag{4.35}$$

Siguiendo los pasos de la demostración del Lema 4.1.1 podemos construir el resultado para $\{z_k\}$, por (4.16) y (4.17)

$$c_i(z_k) = t_k \nabla c_i(x^*)^T w, \; \text{ para todo } \; i \in \mathscr{A}(x^*) \tag{4.36}$$

luego de (4.7), (4.36) y (4.31), se tiene

$$\begin{aligned} \mathscr{L}(z_k, \lambda^*) &= f(z_k) - \sum_{i \in \varepsilon \cup I} \lambda_i^* c_i(z_k) \\ &= f(z_k) - t_k \sum_{i \in \mathscr{A}(x^*)} \lambda_i^* \nabla c_i(x^*)^T w \\ &= f(z_k) \end{aligned} \tag{4.37}$$

A continuación haremos una expansión en serie de Taylor para conseguir una estimación de $\mathscr{L}(z_k, \lambda^*)$ cerca de x^*. Para ello utilizaremos la expresión del Teorema de Taylor y la continuidad de los hessianos $\nabla^2 f$ y $\nabla^2 c_i$, $i \in \varepsilon \cup I$, obteniéndose

$$\mathscr{L}(z_k,\lambda^*) = \mathscr{L}(x^*,\lambda^*) + (z_k - x^*)^T \nabla_x \mathscr{L}(x^*,\lambda^*) + \\ + \frac{1}{2}(z_k - x^*)^T \nabla^2_{xx} \mathscr{L}(x^*,\lambda^*)(z_k - x^*) + o(||z_k - x^*||^2) \quad (4.38)$$

Debido a las condiciones de complementariedad (4.8e), se tiene que $\mathscr{L}(x^*,\lambda^*) = f(x^*)$. De (4.8a), el segundo término del lado derecho es igual a cero. Por lo que, empleando (4.35), se puede reescribir (4.38) de la siguiente forma

$$\mathscr{L}(z_k,\lambda^*) = f(x^*) + \frac{1}{2} t_k^2 w^T \nabla^2_{xx} \mathscr{L}(x^*,\lambda^*) w + o(t_k^2) \quad (4.39)$$

Luego reemplazando (4.37), se obtiene

$$f(z_k) = f(x^*) + \frac{1}{2} t_k^2 w^T \nabla^2_{xx} \mathscr{L}(x^*,\lambda^*) w + o(t_k^2) \quad (4.40)$$

Ahora si $w^T \nabla^2_{xx} \mathscr{L}(x^*,\lambda^*) w < 0$, entonces (4.40) implica que $f(z_k) < f(x^*)$ para todo k suficientemente grande, lo que contradice que x^* sea una solución local. Por lo tanto, la condición (4.33) se cumple. □

Observación: Las condiciones necesarias, suponen que x^* es una solución local y se infieren propiedades de f y c_i, para los índices activos i. Por el contrario las condiciones suficientes, las cuales son condiciones sobre f y c_i, $i \in \varepsilon \cup I$ que son enunciadas a continuación, establecen que x^* es una solución local del problema (4.1), en esta no se requiere la calificación de la restricción, además la desigualdad en (4.33) se reemplaza por una desigualdad estricta.

Teorema 4.1.4 (Condiciones suficientes de segundo orden)**.** *Sea $x^* \in \mathbb{R}^n$, existe un vector multiplicador de Lagrange λ^* de modo que se cumplan las condiciones KKT (4.8). Entonces si*

$$w^T \nabla^2_{xx} \mathscr{L}(x^*,\lambda^*) w > 0, \quad \textit{para todo } w \in \mathscr{C}(x^*,\lambda^*), w \neq 0 \quad (4.41)$$

En consecuencia x^* es una solución local estricta para (4.1). (Nocedal & Wright, 1999)

Demostración.

Afirmamos que existe una vecindad U de x^* y un número $\beta > 0$ tal que

$$f(x) - f(x^*) \geq \beta \|x - x^*\|^2, \qquad \forall x \in \Omega \cap U \tag{4.42}$$

Supongamos que (4.42) no es cierto, es decir:

Para toda vecindad β existe x_k, tal que $f(x_k) - f(x^*) < \frac{1}{k}\|x_k - x^*\|^2$ así existe una sucesión $\{x_k\} \in \Omega - \{x^*\}$ tal que $x_k \to x^*$.

$$f(x_k) - f(x^*) < \frac{1}{k}\|x_k - x^*\|^2, \quad \forall k \tag{4.43}$$

considerando la sucesión limitada $\{\frac{x_k - x^*}{\|x_k - x^*\|}\}$, utilizando el Teorema de Bolzana Weierstrass podemos asumir pasando a una subsucesión si fuese necesario que $\{\frac{x_k - x^*}{\|x_k - x^*\|}\}$ sea convergente y supongamos que $\frac{x_k - x^*}{\|x_k - x^*\|} \to w \in \mathbb{R} - \{0\}$ luego por la Definición 4.1.3 del cono tangente $w \in T_\Omega(x^*) - \{0\}$ y por el Lema 4.1.1 $w \in \mathscr{F}(x^*) - \{0\}$.

Luego utilizando el Teorema de Taylor 2.2.3 y el Corolario 2.2.1 tenemos en (4.43)

$$\frac{1}{k}\|x_k - x^*\|^2 > f(x_k) - f(x^*) = \nabla f(x^*)^T(x_k - x^*) + o(\|x_k - x^*\|)$$

dividiendo esta última relación por $\|x_k - x^*\| > 0$

$$\frac{1}{k}\|x_k - x^*\| > \frac{f(x_k) - f(x^*)}{\|x_k - x^*\|} = \nabla \frac{f(x^*)^T(x_k - x^*)}{\|x_k - x^*\|} + \frac{o(\|x_k - x^*\|)}{\|x_k - x^*\|}$$

Tomando el límite cuando $k \to \infty$ tenemos que

$$\nabla f(x^*)^T w \leq 0 \tag{4.44}$$

y como para λ^* se cumple las condiciones de KKT tenemos en (4.44)

$$\begin{aligned} 0 \geq \nabla f(x^*)^T w &= \sum_{i \in \mathscr{A}(x^*)} \nabla c_i(x^*)^T w \\ &= \sum_{i \in \varepsilon} \lambda_i^* \nabla c_i(x^*)^T w + \sum_{i \in \mathscr{A}(x^*) \cap I} \lambda_i^* \nabla c_i(x^*)^T w \end{aligned}$$

de donde

$$\sum_{i\in\mathscr{A}(x^*)\cap I} \underbrace{\lambda_i^*}_{\geq 0} \nabla \underbrace{c_i(x^*)^T w}_{\geq 0} \leq 0$$

así pues $\nabla c_i(x^*)^T w = 0$, $\forall i \in \mathscr{A}(x^*) \cap I$ con $\lambda_i^* \geq 0$, es decir $w \in \mathscr{C}(x^*, \lambda^*)$.

Por otro lado utilizando el Teorema del Taylor de segundo orden, para todo k se tiene

$$\frac{1}{k}\|x_k - x^*\|^2 \geq f(x_k) - f(x^*) = \nabla f(x^*)^T (x_k - x^*) +$$
$$+\frac{1}{2}(x_k - x^*)^T \nabla^2 f(x^*)(x_k - x^*) + o(\|x - x^*\|^2)$$

Similarmente:

$$0 = c_i(x_k) - c_i(x^*) = \nabla c_i(x^*)^T (x_k - x^*) + \frac{1}{2}(x_k - x^*)^T \nabla^2 f(x^*)(x_k - x^*), \quad \forall i \in \varepsilon \cup (A(x^*) \cap I)$$
$$0 = c_i(x_k) - c_i(x^*) = \nabla c_i(x^*)^T (x_k - x^*) + \frac{1}{2}(x_k - x^*)^T \nabla^2 f(x^*)(x_k - x^*), \quad \forall i \in \mathscr{A}(x^*) \cap I$$

luego multiplicando por λ_i su multiplicador asociado a ambas expresiones y restándolas obtenemos

$$\frac{1}{k}\|x_k - x^*\|^2 > (\nabla c_i(x^*)^T + \sum_{i\in\mathscr{A}(x^*)} \lambda_i \nabla c_i^T)(x_k - x^*) + \frac{1}{2}((x_k - x^*)^T +$$
$$\sum \lambda_i (x_k - x^*)^T \nabla^2 c_i(x^*))(x_k - x^*) + o(\|x_k - x^*\|^2)$$

Lo que contradice (4.43). Concluimos que toda sucesión viable $\{x_k\}$ que se aproxima a x^* satisface $f(x) - f(x^*) \geq \beta\|x - x^*\|^2$ para todo k suficientemente grande, en consecuencia x^* es una solución local estricta. □

4.2. MÉTODO DE PENALIZACIÓN CUADRÁTICA

Un enfoque fundamental para la optimización con restricciones es reemplazar el problema original por una función de penalización que consiste en el objetivo original del problema de optimización mas un término adicional para cada restricción, que es positivo cuando el punto actual x infringe esa restricción y cero en caso contrario.

La mayoría de los métodos definen una sucesión de funciones de penalización, donde los términos de penalización para las violaciones de las restricciones se multiplican por algún factor positivo. Al hacer que este coeficiente sea cada vez mayor, se penaliza las violaciones de restricciones cada vez más severamente, forzando así al minimizador de la función de penalización acercarse cada vez más al conjunto viable del problema con restricciones.

Estos métodos a veces se conocen como métodos de penalización externa, porque el término de penalización para cada restricción es distinto de cero sólo cuando x es no viable con respecto a esta restricción. A menudo, los minimizadores de las funciones de penalización son no viables con respecto al problema original y se acercan a la viabilidad sólo en el límite a medida que el parámetro de penalización se hace cada vez más grande.

La función de penalización más simple de este tipo es la función de penalización cuadrática, en el que los términos de penalización son los cuadrados de las violaciones de las restricciones. Consideremos el problema con restricciones de igualdad para nuestro análisis

$$\min_x f(x) \quad \text{sujeto a} \quad c_i(x) = 0, i \in \varepsilon \tag{4.45}$$

que es un caso especial de (4.1). La función de Penalización Cuadrática $Q(x;\mu)$ para esta formulación es

$$Q(x;\mu) \stackrel{def}{=} f(x) + \frac{1}{2\mu} \sum_{i\in\varepsilon} c_i^2(x) \tag{4.46}$$

donde $\mu > 0$ es el parámetro de penalización. Llevando μ a cero, penalizamos las violaciones de restricciones con una severidad cada vez mayor.

Consideramos una sucesión de valores $\{\mu_k\}$ con $\mu_k \to 0$ mientras que $k \to \infty$, y buscamos el minimizador aproximado x_k de $Q(x;\mu_k)$ para cada k. Debido a que los términos de penalización (4.46) son suaves es decir, cada término c_i^2 tiene almenos tantas derivadas como c_i, podemos usar métodos de optimización sin restricciones para buscar x_k. Los minimizadores aproximados x_k, x_{k-1}, etc se pueden utilizar para identificar un buen punto de partida para la minimización de $Q(\cdot;\mu_{k+1})$ en la iteración $k+1$. Al elegir bien la sucesión $\{\mu_k\}$ y los puntos de partida puede ser posible realizar unos pasos de minimización sin restricciones para cada valor μ_k.

Para el problema general de optimización restringida (4.1), que contiene restricciones de desigualdad así como restricciones de igualdad, podemos definir Q como

$$Q(x;\mu) \stackrel{def}{=} f(x) + \frac{1}{2\mu}\sum_{i\in\varepsilon} c_i^2(x) + \frac{1}{2\mu}\sum_{i\in I}(\text{máx}(-c_i(x),0))^2 \tag{4.47}$$

Un algoritmo general para la función de penalización cuadrática es:

Algoritmo 4.2.1 (Penalización Cuadrática)**.**

Dado $\mu_0 > 0$*, tolerancia* $\tau_0 > 0$*, punto inicial* x_0^s*;*

for $k = 0, 1, 2, \ldots$

Encontrar un minimizador aproximado x_k *de* $Q(\cdot;\mu_k)$*, empezar en* x_k^s*,*

y terminar cuando $\|\nabla Q(x;\mu_k)\| \leq \tau_k$*;*

if *satisface la prueba de convergencia final*

STOP *la solución aproximada* x_k*;*

Elegir un nuevo parámetro de penalización $\mu_{k+1} \in (0,\mu_k)$*;*

Elegir un nuevo punto de inicio x_{k+1}^s*;*

end (for)

4.2.1. Convergencia de la Función de Penalización Cuadrática

Describimos algunas propiedades de convergencia de la función de Penalización Cuadrática en los siguientes dos teoremas.

Teorema 4.2.1. *Supongamos que cada* x_k *es el minimizador global exacto de* $Q(x;\mu_k)$ *en el Algoritmo 4.2.1 y* $\mu_k \to 0$*. Entonces cada punto de acumulación* x^* *de la sucesión* $\{x_k\}$

es una solución del problema (4.45).

Demostración.

Sea $\bar{x}$ una solución global de (4.45), es decir,

$$f(\bar{x}) \leq f(x), \quad \forall x \text{ con } c_i(x) = 0,\ i \in \varepsilon$$

Puesto que x_k minimiza $Q(\cdot;\mu_k)$ para cada k, tenemos que $Q(x_k;\mu_k) \leq Q(\bar{x};\mu_k)$, lo que implica la desigualdad

$$f(x_k) + \frac{1}{2\mu_k}\sum_{i\in\varepsilon} c_i^2(x_k) \leq f(\bar{x}) + \frac{1}{2u_k}\sum_{i\in\varepsilon} c_i^2(\bar{x}) = f(\bar{x}) \tag{4.48}$$

Así obtenemos

$$\sum_{i\in\varepsilon} c_i^2(x_k) \leq 2\mu_k\left[f(\bar{x}) - f(x_k)\right] \tag{4.49}$$

Supongamos que x^* es un punto de acumulación $\{x_k\}$, es decir existe una subsucesión infinita $\mathscr{K}$ tal que

$$\lim_{k\in\mathscr{K}} x_k = x^*$$

Tomando el límite cuando $k \to \infty$, $k \in \mathscr{K}$, en ambos lados de (4.49), obtenemos

$$\sum_{i\in\varepsilon} c_i^2(x^*) = \lim_{k\in\mathscr{K}} \sum_{i\in\varepsilon} c_i^2(x_k) \leq \lim_{k\in\mathscr{K}} 2u_k\left[f(\bar{x}) - f(x_k)\right] = 0$$

Y como $\mu_k \to 0$ tenemos que $c_i(x^*) = 0$ para todo $i \in \varepsilon$, en consecuencia x^* es viable. Además, tomando el límite cuando $k \to \infty$ para $k \in \mathscr{K}$ en (4.48), tenemos por no negatividad de μ_k y de cada $c_i(x_k)^2$ que

$$f(x^*) \leq f(x^*) + \lim_{k\in\mathscr{K}} \frac{1}{2\mu_k}\sum_{i\in\varepsilon} c_i^2(x_k) \leq f(\bar{x})$$

Luego como x^* es un punto viable, $\bar{x}$ es el minimizador global. □

Como este resultado requiere que encontremos el minimizador global para cada subproblema, su propiedad de convergencia para la solución global (4.45) puede ser dificil de realizar en la práctica. El siguiente resulado se refiere a las propiedades de convergencia de la sucesión $\{x_k\}$ cuando permitimos minimizaciones inexactas pero cada vez mas precisas de $Q(\cdot;\mu_k)$. En contraste con el Teorema 4.2.1 muestra que la sucesión es atraída hacia puntos KKT, es decir, puntos que satisfacen la condiciones necesarias de primer orden (4.8), en lugar de un minimizador global. También muestra que las cantidades $c_i(x_k)/\mu_k$ pueden usarse como estimaciones de los multiplicadores de Lagrange λ_i^* en ciertas circunstancias.

Teorema 4.2.2. *Si las tolerancias τ_k en el Algoritmo 4.2.1, satisfacen*

$$\tau_k \to 0$$

y los parámetros de penalización satisfacen $\mu_k \to 0$, entonces para todos los puntos de acumulación x^* de la sucesión $\{x_k\}$ en el que los gradientes de restricción $\nabla c_i(x^*)$ son linealmente independientes, tenemos que x^* es un punto KKT para el problema (4.45). Para tales puntos, tenemos que la subsucesión infinita $\mathcal{K}$ tal que $\lim_{k\in\mathcal{K}} x_k = x^*$ satisface

$$\lim_{k\in K} -\frac{c_i(x_k)}{\mu_k} = \lambda_i^*, \quad \forall i \in \varepsilon \tag{4.50}$$

donde λ^* es un vector multiplicador que satisface las condiciones KKT (4.8).

Demostración.

Diferenciando $Q(x;u_k)$ en (4.46), obtenemos

$$\nabla_x Q(x_k;\mu_k) = \nabla f(x_k) + \sum_{i\in\varepsilon} \frac{c_i(x_k)}{\mu_k} \nabla c_i(x_k) \tag{4.51}$$

Al aplicar el criterio de terminación para el Algoritmo 4.2.1, tenemos que

$$\|\nabla f(x_k) + \sum_{i\in\varepsilon} \frac{c_i(x_k)}{\mu_k} \nabla c_i(x_k)\| \le \tau_k \tag{4.52}$$

y como

$$\|\nabla f(x_k)\| - \left\|\sum_{i\in\varepsilon} \frac{c_i(x_k)}{\mu_k} \nabla c_i(x_k)\right\| \le \left\|\nabla f(x_k) + \sum_{i\in\varepsilon} \frac{c_i(x_k)}{\mu_k} \nabla c_i(x_k)\right\|$$

entonces

$$\left\|\sum_{i\in\varepsilon}\frac{c_i(x_k)}{\mu_k}\nabla c_i(x_k)\right\| \leq \left\|\nabla f(x_k)+\sum_{i\in\varepsilon}\frac{c_i(x_k)}{\mu_k}\nabla c_i(x_k)\right\| + \|\nabla f(x_k)\| \leq \tau_k + \|\nabla f(x_k)\|$$

de donde

$$\|\sum_{i\in\varepsilon} c_i(x_k)\nabla c_i(x_k)\| \leq \mu_k[\tau_k + \|\nabla f(x_k)\|] \tag{4.53}$$

Tomando el límite cuando $k\to\infty$ para $k\in\mathscr{K}$, como el término entre corchetes del lado derecho se aproxima a $\|\nabla f(x^*)\|$, y $\mu_k\to 0$, concluimos que

$$\sum_{i\in\varepsilon} c_i(x^*)\nabla c_i(x^*) = 0 \tag{4.54}$$

Dado que los gradientes de restricción $\nabla c_i(x^*)$ son linealmente independientes, tenemos que $c_i(x^*) = 0$ para todo $i\in\varepsilon$, entonces x^* es viable.

Dado que x^* es viable, se satisface la segunda condición KKT (4.8b). Luego solo necesitamos verificar la primera condición KKT (4.8a) y demostrar que el límite (4.50) se cumple.

Sea $A(x)^T$ la matriz cuyas columnas son las gradientes de las restricciones, es decir,

$$A(x)^T = [\nabla c_i(x)]_{i\in\varepsilon} \tag{4.55}$$

y λ_k denotamos por el vector $-c(x_k)/\mu_k$.

Luego

$$\sum_{i\in\varepsilon}\lambda_k^i\nabla c_i(x_k) = A(x_k)^T\lambda_k$$

y $\nabla_x Q(x_k;\tau_k) = \nabla f(x_k) - \sum_{i\in\varepsilon}\lambda_k\nabla c_i(x_k)$

de donde

$$A(x_k)^T\lambda_k = \nabla f(x_k) - \nabla_x Q(x_k;\mu_k), \qquad \|\nabla_x Q(x_k;\mu)\| \leq \tau_k \tag{4.56}$$

Para todo $k\in\mathscr{K}$ suficientemente grande, como $\nabla c_i(x^*)$ son linealmente independientes, la matriz $A(x_k)$ tiene un rango de columna completo, de modo que $A(x_k)A(x_k)^T$ es no singular. Multiplicando (4.56) por $A(x_k)$, se tiene

$$A(x_k)A(x_k)^T\lambda_k = A(x_k)\left[\nabla f(x_k) - \nabla_x Q(x_k;\mu_k)\right]$$

de donde

$$\lambda_k = \left[A(x_k)A(x_k)^T\right]^{-1} A(x_k)\left[\nabla f(x_k) - \nabla_x Q(x_k;\mu_k)\right]$$

Tomando el límite cuando $k \to \infty$ en $\mathscr{K}$ y considerando que $\tau_k \to 0$, tenemos que

$$\lim_{k\in K} \lambda_k = \left[A(x^*)A(x^*)^T\right]^{-1} A(x^*)\nabla f(x^*)$$

donde el límite lo denotamos por λ^*.

Y como $\lambda^* = \lim\limits_{k\in\mathscr{K}} -\dfrac{c_i(x_k)}{\mu_k}$.

Además de (4.52) se tiene

$$\left\|\nabla f(x_k) - \sum_{i\in\varepsilon} \lambda_k \nabla c_i(x_k)\right\| \le \tau_k$$

entonces tomando el límite, concluimos que

$$\nabla f(x^*) - A(x^*)^T\lambda^* = \nabla f(x^*) - \sum_{i\in\varepsilon} \lambda_i^* \nabla c_i(x^*) = 0$$

De modo que λ^* satisface la primera condición KKT (4.8a). Así x^* es un punto KKT, con su multiplicador de Lagrange λ^*. $\square$

4.3. MÉTODO LAGRANGIANO AUMENTADO

El Método Lagrangiano Aumentado es un método de optimización restringido para obtener el mínimo de una función sujeta a restricciones de igualdad y desigualdad como (4.1).

El método transforma el problema de minimización restringida en un problema de minimización no restringida restando una estimación de cada multiplicador de Lagrange multiplicado por la función de restricción y agregando a la función objetivo un término que aumenta con la magnitud de la violación de la restricción.

El Método Lagrangiano Aumentado conocido también como método de los multiplicadores, reduce la posibilidad de un mal condicionamiento introduciendo estimaciones explícitas del multiplicador de Lagrange en la función a minimizar que se conoce como función Lagrangiana Aumentada.

Este método está relacionado con el Método de Penalización Cuadrática pero reduce la posibilidad de un mal condicionamiento de los subproblemas que se generan en este enfoque al introducir estimaciones explícitas del multiplicador de Lagrange en cada paso en la función que se va a minimizar. Prescinde de la necesidad de iteraciones para seguir siendo estrictamente viables con respecto a las restricciones de desigualdad.

En esta sección utilizamos superíndices k y $k+1$ en las estimaciones del multiplicador de Lagrange para denotar el índice de iteración y los subíndices i para denotar los índices componentes del vector λ.

4.3.1. Marco de Motivación y Algoritmo

Primero consideramos el problema con igualad de restricciones (4.45). La función de penalización cuadrática dada por (4.46) penaliza las violaciones de restricciones elevando al cuadrado las inviabilidades y escalándolas en $1/(2\mu)$. Sin embargo, como vemos en el Teorema 4.2.2, los minimizadores aproximados x_k de $Q(x;\mu_k)$ no satisfacen del todo las condiciones de viabilidad $c_i(x)=0,\ i\in\varepsilon$. En cambio, se perturban ligeramente como vemos en (4.50) para satisfacer aproximadamente

$$c_i(x_k) = -\mu_k\lambda_i^*, \quad \forall i\in\varepsilon \tag{4.57}$$

Esta perturbación desaparece a medida que $u_k \to 0$, pero nos podemos preguntar si se puede alterar la función $Q(x;\mu_k)$ para evitar esta perturbación sistemática, es decir, hacer que los minimizadores aproximados satisfagan mejor las restricciones de igualdad $c_i(x) = 0$. Al hacerlo, podemos evitar la necesidad de disminuir μ a cero, y así evitar el mal condicionamiento y los problemas numéricos asociados con $Q(x;\mu)$ para valores pequeños de este parámetro de penalización.

La función Lagrangiana Aumentada $\mathscr{L}_A(x,\lambda;\mu)$ logra estos objetivos al incluir una estimación explícita de los multiplicadores de Lagrange λ, con base en la fórmula (4.50), en el objetivo. De la definición

$$\mathscr{L}_A(x,\lambda;\mu) \overset{def}{=} f(x) - \sum_{i\in\varepsilon} \lambda_i c_i(x) + \frac{1}{2\mu}\sum_{i\in\varepsilon} c_i^2(x) \tag{4.58}$$

Vemos que el Lagrangiano Aumentado difiere del Lagrangiano estándar (4.7) para (4.45) por la presencia de los términos al cuadrado, mientras que difiere de la función de Penalización Cuadrática (4.46) con la presencia del término de suma que involucra los λ. En ese sentido, es una combinación de las funciones de penalización Lagrangiana y Cuadrática. Cuando diferenciamos con respecto a x, obtenemos

$$\nabla_x \mathscr{L}_A(x,\lambda;\mu) = \nabla f(x) - \sum_{i\in\varepsilon} [\lambda_i - c_i(x)/\mu] \nabla c_i(x) \tag{4.59}$$

Usando x_k para denotar el minimizador aproximado de $\mathscr{L}_A(x,\lambda;\mu_k)$, podemos usar la lógica de la demostración del Teorema (4.2.2), para deducir que

$$\lambda_i^* \approx \lambda_i^k - c_i(x_k)/u_k, \quad \forall i \in \varepsilon \tag{4.60}$$

Reordenando esta expresión tenemos

$$c_i(x_k) \approx -\mu_k\left(\lambda_i^* - \lambda_i^k\right), \quad \forall i \in \varepsilon$$

Entonces concluimos que si λ^k está cerca del vector multiplicador óptimo λ^*, la no viabilidad en x_k será mucho más pequeña que μ_k, en lugar de ser proporcional a μ_k como en (4.57).

¿Cómo podemos actualizar las estimaciones del multiplicador λ^k de iteración en iteración, de modo que se aproxime a λ^* cada vez con mayor presición, basándose en la información actual?. La ecuación (4.60) sugiere inmediatamente la fórmula

$$\lambda_i^{k+1} = \lambda_i^k - c_i(x_k)/\mu_k, \quad \forall i \in \varepsilon \tag{4.61}$$

Algoritmo 4.3.1 (Método de Multiplicadores - Restricciones de igualdad)**.**

Dado $\mu_0 > 0$*, tolerancia* $\tau_0 > 0$*, puntos de partida* x_0^s *y* λ^0*;*

for $k = 0, 1, 2, \ldots$

Encontrar un minimizador aproximado x_k *de* $\mathcal{L}_A\left(\cdot, \lambda^k; \mu_k\right)$*, empezar en* x_k^s*,*

y terminar cuando $\left\|\nabla_x \mathcal{L}_A\left(x, \lambda^k; \mu_k\right)\right\| \leq \tau_k$*;*

if *satisface la prueba de convergencia final*

STOP *con la solución aproximada* x_k*;*

Actualizar los multiplicadores del Lagrange usando (4.61) para obtener λ^{k+1}*;*

Elegir un nuevo parámetro de penalización $\mu_{k+1} \in (0, \mu_k)$*;*

Establecer el punto de partida para la siguiente iteración en $x_{k+1}^s = x_k$*;*

end (for)

4.3.2. Lagrangiano Aumentado para Restricciones de Igualdad

Ahora demostramos dos resultado que justifican el uso de la función Lagrangiana Aumentada y el método de los multiplicadores. Para simplificar, limitamos nuestra atención al caso con restricciones de igualdad, es decir, el problema (4.45) para el cual el Lagrangiano Aumentado está dado por (4.58).

El primer resultado valida el enfoque del Algoritmo 4.3.1 al mostrar que cuando conocemos el vector multiplicador de Lagrange exacto λ^*, la solución x^* de (4.45) es un minimizador estricto de $\mathcal{L}_A(x, \lambda^*; \mu)$ para todo μ suficientemente pequeño. Aunque no conocemos λ^* exactamente en la práctica, el resultado y su prueba sugieren fuertemente que podemos obtener una buena estimación de x^* minimizando $\mathscr{L}_A(x,\lambda^*;\mu)$ incluso cuando μ es particularmente cercano a cero, siempre que λ sea una estimación razonable de λ^*.

Teorema 4.3.1. *Sea x^* una solución local de (4.1) en la que se satisface la LICQ (es decir, los gradientes $\nabla c_i(x^*), i \in \varepsilon$, son vectores linealmente independientes), y se satisfacen las condiciones suficientes de segundo orden especificadas en el Teorema 4.1.3 para $\lambda = \lambda^*$. Entonces existe un valor $\overline{\mu}$ tal que para todo $\mu \in (0,\overline{\mu}]$, x^* es un minimizador local estricto de $\mathscr{L}_A(x,\lambda^*;\mu)$*

Demostración.

Probamos el resultado demostrando que x^* satisface las condiciones suficientes de segundo orden para ser un minimizador local estricto de $\mathscr{L}(x,\lambda^*;\mu)$, es decir

$$\nabla_x \mathscr{L}_A(x^*,\lambda^*;\mu) = 0, \quad \nabla^2_{xx}\mathscr{L}_A(x^*,\lambda^*;\mu) \text{ definido positivo} \tag{4.62}$$

Usando las condiciones KKT (4.8) y la fórmula (4.59), tenemos que

$$\begin{aligned}\nabla_x \mathscr{L}_A(x^*,\lambda^*;\mu) &= \nabla f(x^*) - \sum_{i\in\varepsilon}[\lambda_i^* - c_i(x^*)/\mu]\nabla c_i(x^*)\\ &= \nabla f(x^*) - \sum_{i\in\varepsilon}\lambda_i^*\nabla c_i(x^*) = \nabla_x \mathscr{L}(x^*,\lambda^*) = 0\end{aligned}$$

verificando la primera parte de (4.62) indenpendientemente de μ.

Ahora demostramos la segunda parte de (4.62) demostrando que

$$u^T\nabla^2_{xx}\mathscr{L}_A(x^*,\lambda^*;\mu)u > 0$$

para todo $u \in \mathbb{R}^n$ y todo $\mu > 0$ suficientemente pequeño. De (4.55), tenemos

$$A^T = [\nabla c_i^*]\, i \in \varepsilon$$

y notamos que por LICQ de la Definición 4.1.5, A tiene rango de fila completo. Además

$$\nabla_{xx}^2 \mathcal{L}_A(x^*, \lambda^*; \mu) = \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) + \frac{1}{\mu} A^T A \tag{4.63}$$

Según el teorema fundamental del álgebra, podemos dividir cualquier $u \in \mathbb{R}^n$ en componentes $NullA$ (espacio nulo de A) y $RangoA^T$, y escribir

$$u = w + A^T v,$$

donde $w \in NullA$ y v es un vector en $\mathbb{R}^{|\varepsilon|}$. Usando (4.63) y las propiedades de w y v, tenemos que

$$\begin{aligned} u^T \nabla_{xx}^2 \mathcal{L}_A(x^*, \lambda^*; \mu) u = w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w + 2w^T \nabla_{xx}^2 w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) A^T v \\ + v^T A \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) A^T v + v^T A \left((1/\mu) A^T A\right) A^T v \end{aligned} \tag{4.64}$$

Buscamos límites en los tres términos del lado derecho de esta expresión.

Debido a (4.41) y la compacidad de la esfera unitaria intersectada con $NullA$, existe un escalar $a > 0$ tal que

$$w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) w \geq a \|w\|^2, \quad \forall w \in NullA$$

Dando un límite inferior en el primer término del lado derecho en (4.64). Para el segundo término, definimos $b \stackrel{def}{=} \left\|\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) A^T\right\|$, de modo que

$$2w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) A^T v \geq -2b \|w\| \|v\|$$

Dando un límite inferior para el segundo término. Para el tercer término, si definimos $c \stackrel{def}{=} \left\|A \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) A^T\right\|$, obtenemos

$$v^T A \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) A^T v \geq -c \|v\|^2$$

Finalmente, si d es el valor propio más pequeño de AA^T, tenemos

$$v^T A \left(\frac{1}{\mu} A^T A\right) A^T v \geq \frac{1}{\mu} \left\|AA^T v\right\|^2 \geq \frac{d^2}{\mu} \|v\|^2$$

Sustituyendo estos límites inferiores en (4.64), obtenemos

$$\begin{aligned} u^T \nabla^2_{xx} \mathcal{L}_A(x^*, \lambda^*, \mu) u &\geq a\|w\|^2 - 2b\|v\|\|w\| + \left(d^2/\mu - c\right)\|v\|^2 \\ &= a\left[\|w\| - (b/a)\|v\|\right]^2 + \left(d^2/\mu - c - b^2/a\right)\|v\|^2 \end{aligned} \tag{4.65}$$

El primer término del lado derecho de esta expresión es claramente no negativo. Ya que $d > 0$ por el rango completo de A, el segundo término también es no negativo, siempre que $\overline{\mu}$ sea cualquier valor tal que

$$\frac{d^2}{\overline{\mu}} - c - \frac{b^2}{a} > 0$$

donde $\mu \in \langle 0, \overline{\mu}]$. En efecto, el lado derecho de (4.65) es estrictamente positivo a menos que $v = 0$ y $w = 0$, lo que implica que $\nabla^2_{xx}\mathcal{L}_A(x^*, \lambda^*; \mu)$ es definido positivo, según sea necesario. Por lo tanto, hemos verificado (4.62) y es completada la demostración.

□

4.3.3. Lagrangiano Aumentado para Restricciones de Igualdad y Desigualdad

Vamos a transformar todo problema con restricciones de igualdad y desigualdad (4.1) a un problema con restricciones de igualdad introduciendo la variable de holgura t_i de la siguiente forma:

$$\min_x f(x) \text{ sujeto a } \begin{cases} c_i(x) = 0, & i \in \varepsilon \\ -c_i(x) + t_i^2 = 0, & i \in I \end{cases} \tag{4.66}$$

Para todo $\mu > 0$ definimos el Lagrangiano aumentado para el problema (4.66)

$$\begin{aligned} \widetilde{\mathcal{L}_A}((x,t), \lambda; \mu) = f(x) + \sum_{i \in \varepsilon} \lambda_i c_i(x) + \sum_{i \in I} \lambda_i \left(-c_i(x) + t_i^2\right) + \\ + \frac{1}{2\mu}\left(\sum_{i \in \varepsilon} c_i^2(x) + \sum_{i \in I}\left(-c_i(x) + t_i^2\right)^2\right) \end{aligned} \tag{4.67}$$

la minimización

$$\min \widetilde{\mathcal{L}_A}((x,t), \lambda; c) \tag{4.68}$$

en relación a t puede ser hecha explícitamente. Tenemos que

$$\begin{aligned}\frac{\partial \widetilde{\mathscr{L}_A}}{\partial t_i}\left((x,t),\lambda;\mu\right) &= 2\lambda_i t_i + \frac{2}{2\mu}\left(-c_i\left(x\right)+t_i^2\right)\left(2t_i\right)=0\\ &= 2t_i\left[\lambda_i+\frac{1}{\mu}\left(-c_i\left(x\right)+t_i^2\right)\right]=0\end{aligned}$$

luego

$$t_i=0 \quad \text{o} \quad \left[\lambda_i+\frac{1}{\mu}\left(-c_i\left(x\right)+t_i^2\right)\right]=0$$

Así, para todo $x\in\mathbb{R}^n$ fijo, el mínimo global se alcanza en el punto $t_i\left(x,\lambda;\mu\right)$.

$$t_i\left(x,\lambda;\mu\right)=\begin{cases}0, \text{ si } -\mu\lambda_i+c_i\left(x\right)<0, \quad \left(c_i\left(x\right)<\mu\lambda_i\right)\\ \sqrt{-\mu\lambda_i+c_i\left(x\right)}=\sqrt{c_i\left(x\right)-\mu\lambda_i}, \text{ caso contrario}\end{cases} \tag{4.69}$$

Además:

$$\begin{aligned}-c_i\left(x\right)+t_i^2 &= \begin{cases}-c_i\left(x\right), \text{ si } c_i\left(x\right)<\mu\lambda_i\\ -c_i\left(x\right)+\left(-\mu\lambda_i+c_i\left(x\right)\right), \text{ si } c_i\left(x\right)\geq\mu\lambda_i\end{cases}\\ &= \begin{cases}-c_i\left(x\right), \text{ si } c_i\left(x\right)<\mu\lambda_i, \quad \left(-\mu\lambda_i<-c_i\left(x\right)\right)\\ -\mu\lambda_i, \text{ si } c_i\left(x\right)\geq\mu\lambda_i, \quad \left(-\mu\lambda_i\geq-c_i\left(x\right)\right)\end{cases}\end{aligned}$$

luego

$$-c_i\left(x\right)+t_i^2=\text{máx}\left\{-c_i\left(x\right),-\mu\lambda_i\right\} \tag{4.70}$$

Así el Lagrangiano aumentado de (4.67) queda como:

$$\begin{aligned}\widetilde{\mathscr{L}_A}\left(x,\lambda;\mu\right) &= f\left(x\right)+\sum_{i\in\varepsilon}\lambda_i c_i\left(x\right)+\sum_{i\in I}\lambda_i\,\text{máx}\left\{-c_i\left(x\right),-\mu\lambda_i\right\}+\\ &\quad+\frac{1}{2\mu}\left[\sum_{i\in\varepsilon}c_i^2\left(x\right)+\sum_{i\in I}\left(\text{máx}\left\{-c_i\left(x\right),-\mu\lambda_i\right\}\right)^2\right]\\ &= f\left(x\right)+\sum_{i\in\varepsilon}\lambda_i c_i\left(x\right)+\frac{1}{2\mu}\sum_{i\in\varepsilon}c_i^2\left(x\right)+\sum_{i\in I}\lambda_i\,\text{máx}\left\{-c_i\left(x\right),-\mu\lambda_i\right\}+\\ &\quad+\frac{1}{2\mu}\sum_{i\in I}\left(\text{máx}\left\{-c_i\left(x\right),-\mu\lambda_i\right\}\right)^2\end{aligned} \tag{4.71}$$

Por otro lado sea

$$w_i = \lambda_i \text{máx}\left\{-c_i(x), -\mu\lambda_i\right\} + \frac{1}{2\mu}\text{máx}\left(\left\{-c_i(x), -\mu\lambda_i\right\}\right)^2 \tag{4.72}$$

entonces, si $-c_i(x) \geq -\mu\lambda_i$, $\left(\lambda_i - \frac{c_i(x)}{\mu} \geq 0\right)$, se tiene

$$\begin{aligned}
w_i &= -\lambda_i c_i(x) + \frac{1}{2\mu}c_i^2(x) \\
&= \frac{\mu}{2}\left[\lambda_i^2 - \frac{2\lambda_i c_i(x)}{\mu} + \frac{c_i^2(x)}{\mu^2}\right] - \lambda_i^2 \\
&= \frac{\mu}{2}\left[\lambda_i - \frac{c_i(x)}{\mu}\right]^2 - \lambda_i^2 \\
&= \frac{\mu}{2}\left[\text{máx}\left\{0, \lambda_i - \frac{c_i(x)}{\mu}\right\}\right]^2 - \lambda_i^2
\end{aligned}$$

en cambio si $-c_i(x) < -\mu\lambda_i$ $\left(\lambda_i - \frac{c_i(x)}{\mu} < 0\right)$, se tiene

$$\begin{aligned}
w_i &= -\mu\lambda_i^2 + \frac{\mu^2\lambda_i^2}{2\mu} \\
&= -\frac{\mu}{2}\lambda_i^2 \\
&= \frac{\mu}{2}\left[\text{máx}\left\{0, \lambda_i - \frac{c_i(x)}{\mu}\right\}\right]^2 - \lambda_i^2
\end{aligned}$$

Por consiguiente

$$w_i = \frac{\mu}{2}\left(\text{máx}\left\{0, \lambda_i - \frac{c_i(x)}{\mu}\right\}\right)^2 - \lambda_i^2 \tag{4.73}$$

así pues (4.71) se reduce a:

$$\begin{aligned}
\mathscr{L}_A(x, \lambda; \mu) = f(x) + \sum_{i\in\varepsilon}\lambda_i c_i(x) + \frac{1}{2\mu}\sum_{i\in\varepsilon}c_i^2(x) + \\
+ \frac{\mu}{2}\left[\sum_{i\in I}\left(\text{máx}\left\{0, \lambda_i - \frac{c_i(x)}{\mu}\right\}\right)^2 - \lambda_i^2\right]
\end{aligned} \tag{4.74}$$

Por tanto, en vez de (4.68) consideraremos el problema

$$\text{mín}\,\mathscr{L}_A(x, \lambda; \mu), \quad x \in \mathbb{R}^n \tag{4.75}$$

El teorema siguiente muestra que los puntos KKT del problema original (4.1) corresponden a puntos estacionarios sin restricciones del Lagrangiano Aumentado de este problema, para cualquier $\mu > 0$. Antes precisaremos un resultado auxiliar.

Lema 4.3.1. *Sean $a \in \mathbb{R}$ y $b \in \mathbb{R}$. Entonces*

$$a = \text{máx}\{0, b\} \tag{4.76}$$

si, y solamente si,

$$a - b \geq 0,\ a(a-b) = 0,\ a \geq 0 \tag{4.77}$$

Demostración.

Observamos que (4.76) significa que

$$a \geq 0,\ a \geq b \text{ y } a = 0 \text{ o } a = b$$

estas condiciones también pueden ser escritas como (4.77). □

Teorema 4.3.2 (Puntos estacionarios del Lagrangiano aumentado)**.** *Sea $f, c_i : \mathbb{R}^n \to \mathbb{R}$ con $i \in \varepsilon \cup I$, funciones diferenciables en el punto $\overline{x} \in \mathbb{R}^n$. Entonces para $\mu > 0$ cualquiera, el punto $\left(\overline{x}, \overline{\lambda}\right)$ es una solución del sistema de Karush Kuhn Tucker del problema (2.2) si, y solamente si,*

$$\nabla \mathscr{L}_A\left(\overline{x}, \overline{\lambda}; \mu\right) = 0$$

Demostración.

Se tiene que:

(a)

$$\frac{\partial \mathscr{L}_A}{\partial x}\left(\overline{x}, \overline{\lambda}; \mu\right) = \frac{\partial f(\overline{x})}{\partial x} + \sum_{i\in\varepsilon} \overline{\lambda_i} \frac{\partial c_i(\overline{x})}{\partial x} + \frac{1}{\mu} \sum_{i\in\varepsilon} c_i(\overline{x}) \frac{\partial c_i(\overline{x})}{\partial x} + \\ + \mu \sum_{i\in I}\left[\text{máx}\left\{0, \overline{\lambda_i} - \frac{c_i(\overline{x})}{\mu}\right\}\right]\left(\frac{1}{\mu}\frac{\partial c_i(\overline{x})}{\partial x}\right)$$

luego:

$$\nabla_x \mathscr{L}_{\mathscr{A}}\left(\overline{x}, \overline{\lambda}; \mu\right) = \nabla_x f(\overline{x}) + \sum_{i\in\varepsilon}\left(\overline{\lambda_i}\nabla_x c_i(\overline{x}) + \frac{c_i(\overline{x})}{\mu}\nabla_x c_i(\overline{x})\right) + \\ + \sum_{i\in I}\left[\text{máx}\left\{0, \overline{\lambda_i} - \frac{c_i(\overline{x})}{\mu}\right\}\right]\nabla_x c_i(\overline{x}) \tag{4.78}$$

(b) Si $i \in \varepsilon$

$$\frac{\partial \mathscr{L}_A}{\partial \lambda_i}\left(\overline{x},\overline{\lambda};\mu\right) = c_i\left(\overline{x}\right) \text{ y como } c_i\left(x\right) = 0 \text{ para } i \in \varepsilon$$

en particular

$$\frac{\partial \mathscr{L}_A}{\partial \lambda_i}\left(\overline{x},\overline{\lambda};\mu\right) = 0 \text{ si y solo si } c_i\left(\overline{x}\right) = 0, \forall i \in \varepsilon \tag{4.79}$$

(c) Si $i \in I$

$$\frac{\partial \mathscr{L}_A}{\partial \lambda_i}\left(\overline{x},\overline{\lambda};\mu\right) = \left(\text{máx}\left\{0, \overline{\lambda_i} - \frac{c_i\left(\overline{x}\right)}{\mu}\right\} - \overline{\lambda_i}\right)$$

en particular

$$\frac{\partial \mathscr{L}_A}{\lambda_i}\left(\overline{x},\overline{\lambda},\mu\right) = 0 \tag{4.80}$$

si y solo si

$$\overline{\lambda}_i = \text{máx}\left\{0, \overline{\lambda}_i, -\frac{c_i\left(\overline{x}\right)}{\mu}\right\}, \quad \forall i \in I \tag{4.81}$$

de donde por el Lema 4.3.1

$$\frac{c_i\left(\overline{x}\right)}{\mu} \geq 0, \quad \frac{\overline{\lambda}_i c_i\left(\overline{x}\right)}{\mu} = 0, \quad \lambda_i \geq 0$$

es decir que:

$$c_i\left(\overline{x}\right) \geq 0, \quad \lambda_i c_i\left(\overline{x}\right) = 0, \quad \overline{\lambda}_i \geq 0$$

Luego considerando (4.79) y (4.81) en (4.78), tenemos que

$$\nabla_x \mathscr{L}_A\left(\overline{x},\overline{\lambda};\mu\right) = \nabla_x f\left(\overline{x}\right) + \sum_{i\in\varepsilon} \overline{\lambda}_i \nabla_x c_i\left(\overline{x}\right) + \sum_{i\in I} \overline{\lambda}_i \nabla_x c_i\left(\overline{x}\right) \tag{4.82}$$

Concluyéndose de (4.82), (4.79) y (4.81) que

$$\nabla \mathscr{L}_A\left(\overline{x},\overline{\lambda};\mu\right) = 0$$

$\square$

Los siguientes lemas garantizan que los puntos estacionarios del problema con variables de holgura (4.66) genera puntos estacionarios del problema original.

Lema 4.3.2. *Sea $f, c_i : \mathbb{R}^n \to \mathbb{R}$ diferenciables en el punto $\tilde{x} \in \mathbb{R}^n$, entonces para todo $\mu > 0$ y cualquier λ, $\tilde{x}$ es un punto estacionario del problema (4.75) con la función objetivo definida por la fórmula (4.74), si y solamente si, el punto $(\tilde{x}, t(\tilde{x}))$, donde $t(\tilde{x}) \in \mathbb{R}^m$ está definido en (4.69), es estacionario para el problema (4.68) con función objetivo dada por (4.67).*

Demostración.

Sea $\tilde{x}$ un punto estacionario del problema (4.75) con función objetivo definida por (4.74), es decir $\nabla_x \mathscr{L}_A(x, \lambda; \mu) = 0$.

Considerando el problema (4.68) con la función objetivo dada por (4.67), se tiene

$$\begin{aligned}\nabla_x \widetilde{\mathscr{L}_A}((x,t), \lambda, \mu) = \nabla_x f(x) + \sum_{i\in\varepsilon} \lambda_i \nabla_x c_i(x) - \sum_{i\in I} \lambda_i \nabla_x c_i(x) + \\ + \frac{1}{\mu}\left(\sum_{i\in\varepsilon} c_i(x) \nabla_x c_i(x)\right) - \frac{1}{\mu}\left(\sum_{i\in I}\left(-c_i(x) + t_i^2\right)\nabla_x c_i(x)\right)\end{aligned}$$

$$\begin{aligned}&= \nabla_x f(x) + \sum_{i\in\varepsilon}\left(\lambda_i + \frac{c_i(x)}{\mu}\right)\nabla_x c_i(x) - \sum_{i\in I}\left(\lambda_i + \frac{-c_i(x) + t_i^2}{\mu}\right)\nabla_x c_i(x) \\ &= \nabla_x f(x) + \sum_{i\in\varepsilon}\left(\lambda_i + \frac{c_i(x)}{\mu}\right)\nabla_x c_i(x) - \frac{1}{\mu}\sum_{i\in I}\left(\mu\lambda_i - c_i(x) + t_i^2\right)\nabla_x c_i(x)\end{aligned}$$

Por otro lado, utilizando (4.70) se tiene que

$$\begin{aligned}\mu\lambda_i - c_i(x) + t_i^2 &= \mu\lambda_i + \text{máx}\left\{-c_i(\tilde{x}), -\mu\lambda_i\right\} \\ &= \mu\lambda_i + \begin{cases} -c_i(x), & \text{si } -c_i(x) \geq -\mu\lambda_i \\ -\mu\lambda_i, & \text{si } -c_i(x) < -\mu\lambda_i \end{cases} \\ &= \begin{cases} \mu\lambda_i - c_i(x), & \text{si } \mu\lambda_i - c_i(x) \geq 0 \\ 0, & \text{si } \mu\lambda_i - c_i(x) < 0 \end{cases} \\ &= \text{máx}\left\{0, \mu\lambda_i - c_i(x)\right\}\end{aligned}$$

de modo que

$$\begin{aligned}\nabla_x \widetilde{\mathscr{L}_A}\left(\left(x,t\right),\lambda;\mu\right) &= \nabla_x f\left(x\right) + \sum_{i\in\varepsilon}\left(\lambda_i + \frac{c_i\left(x\right)}{\mu}\right)\nabla_x c_i\left(x\right) - \frac{1}{\mu}\sum_{i\in I}\left(\text{máx}\left\{0,\mu\lambda_i - c_i\left(x\right)\right\}\right)\nabla_x c_i\left(x\right)\\ &= \nabla_x f\left(x\right) + \sum_{i\in\varepsilon}\left(\lambda_i + \frac{c_i\left(x\right)}{\mu}\right)\nabla_x c_i\left(x\right) - \sum_{i\in I}\left(\text{máx}\left\{0,-\lambda_i + \frac{c_i\left(x\right)}{\mu}\right\}\right)\nabla_x c_i\left(x\right)\end{aligned}$$

así

$$\nabla_x \widetilde{\mathscr{L}_A}\left(\left(x,t\right),\lambda;\mu\right) = \nabla_x \mathscr{L}_A\left(x,\lambda;\mu\right)$$

y si $\tilde{x}$ es un punto estacionario del problema (4.41) con función objetivo dada por (4.40), entonces

$$\nabla_x \widetilde{\mathscr{L}_A}\left(\left(\tilde{x},t\left(\tilde{x}\right)\right),\lambda;\mu\right) = 0 = \nabla_x \mathscr{L}_A\left(\tilde{x},\lambda;\mu\right)$$

Por otro lado,

$$\begin{aligned}\frac{\partial \widetilde{\mathscr{L}_A}}{\partial t_i}\left(\left(x,t\right),\lambda;\mu\right) &= 2\lambda_i t_i + \frac{2}{2\mu}\left(-c_i\left(x\right)+t_i^2\right)\left(2t_i\right)\\ &= \frac{2}{\mu}t_i\left(\mu\lambda_i - c_i\left(x\right)+t_i^2\right)\end{aligned}$$

y en $t_i = t_i\left(\widetilde{x}\right)$

$$\nabla_t \frac{\partial \widetilde{\mathscr{L}_A}}{\partial t_i}\left(\left(\tilde{x},t\left(\tilde{x}\right)\right),\lambda;\mu\right) = 0$$

$\square$

Lema 4.3.3. *Para $f,c_i:\mathbb{R}^n\to\mathbb{R}$ cualquiera, si para algún $\mu>0$ el punto $\left(\tilde{x},\tilde{t}\right)$ es una solución local del problema (4.68) con función objetivo definida por la fórmula (4.67), entonces $\tilde{t}$ coincide (ignorando los signos de las coordenadas) con el punto $t\left(\tilde{x}\right)$, definido de acuerdo con (4.69).*

Demostración.

Definimos la función

$$\begin{aligned}\widehat{\mathscr{L}_A}\left(\left(\tilde{x},t\right),\lambda;\mu\right) = f\left(\tilde{x}\right) + \sum_{i\in\varepsilon}\lambda_i c_i\left(\tilde{x}\right) + \sum_{i\in I}\lambda_i\left(-c_i\left(\tilde{x}\right)+t_i^2\right) + \\ + \frac{1}{2\mu}\left(\sum_{i\in\varepsilon}c_i^2\left(\tilde{x}\right) + \sum_{i\in I}\left(-c_i\left(\tilde{x}\right)+t_i^2\right)^2\right)\end{aligned}$$

luego,

$$\begin{aligned}\frac{\partial \widehat{\mathscr{L}_A}}{\partial t_i}\left(\left(\tilde{x},t\right),\lambda;\mu\right) &= 2\lambda_i t_i + \frac{2}{2\mu}\left(-c_i\left(\tilde{x}\right)+t_i^2\right)\left(2t_i\right)\\ &= 2t_i\left[\lambda_i + \frac{1}{\mu}\left(-c_i\left(\tilde{x}\right)+t_i^2\right)\right]\\ &= \frac{2}{\mu}t_i\left(\mu\lambda_i - c_i\left(\tilde{x}\right)+t_i^2\right)\end{aligned}$$

entonces

$$\frac{2\tilde{t}_i}{\mu}\left(\mu\lambda_i + \left(-c_i\left(x\right)+\tilde{t}_i^2\right)\right) = 0$$

de donde

$$\tilde{t}_i = 0 \ \text{ o } \ \left[\lambda_i\mu + \left(-c_i\left(\tilde{x}\right)+\tilde{t}_i^2\right)\right] = 0 \tag{4.83}$$

si

$$\mu\lambda_i - c_i\left(\tilde{x}\right) \geq 0$$

se tiene que

$$\tilde{t}_i = 0 = t_i\left(\tilde{x}\right) \tag{4.84}$$

y por otro lado si $\mu\lambda_i - c_i\left(\tilde{x}\right) < 0$ supongamos que $\tilde{t}_i^2 \neq \left(t_i\left(\tilde{x}\right)\right)^2 = c_i\left(\tilde{x}\right) - \mu\lambda_i$ y luego por (4.83) se tiene que $\tilde{t}_i = 0$. Por otro lado

$$\begin{aligned}\frac{\partial^2 \widehat{\mathscr{L}_A}}{\partial t_i^2}\left(\left(\tilde{x},\tilde{t}\right),\lambda;\mu\right) &= \frac{2}{\mu}\left(\mu\lambda_i - c_i\left(\tilde{x}\right)+3\tilde{t}_i^2\right)\\ &= \frac{2}{\mu}\left(\mu\lambda_i - c_i\left(\tilde{x}\right)\right) < 0\end{aligned}$$

Puesto que $\tilde{t}$ es un minimizador local y por el Teorema 4.1.3 se tiene $\nabla^2\widehat{\mathscr{L}_A}\left(\left(\tilde{x},\tilde{t}\right),\lambda;\mu\right) \geq 0$ y como esta matriz es diagonal, entonces $\dfrac{\partial^2 \widehat{\mathscr{L}_A}}{\partial t_i^2}\left(\left(\tilde{x},\tilde{t}\right),\lambda;\mu\right) \geq 0$

Así se tiene que

$$\tilde{t}_i = \pm t_i\left(\tilde{x}\right).$$

□

CAPÍTULO V

CONCLUSIONES

Las conclusiones del presente trabajo de investigación son las siguientes:

1. Si x^* es una solución local del problema

$$\min_x f(x) \text{ sujeto a } \begin{cases} c_i(x) = 0, & i \in \varepsilon \\ c_i(x) \geq 0, & i \in I \end{cases} \tag{\#}$$

   y satisface la condición LICQ, entonces existe un vector multiplicador de Lagrange λ^* tal que se satisface las condicones de Karush Kuhn Tucker (KKT):

$$\begin{aligned} \nabla_x \mathscr{L}(x^*, \lambda^*) &= 0 \\ c_i(x^*) &= 0, \ \forall i \in \varepsilon \\ c_i(x^*) &\geq 0, \ \forall i \in I \\ \lambda_i^* &\geq 0, \ \forall i \in I \\ \lambda_i^* c_i(x^*) &= 0, \ \forall i \in \varepsilon \cup I \end{aligned}$$

   donde $\mathscr{L}(x, \lambda) = f(x) - \sum_{i \in E \cup I} \lambda_i c_i(x)$

2. Sea x^* una solución local del problema (#) que satisface las condición LICQ; si λ^* es un vector multiplicador de Lagrange que cumple las condiciones (KKT). Entonces se cumple

$$w^T \nabla_{xx}^2 \mathscr{L}(x^*, \lambda^*) w \geq 0, \ \forall w \in \mathscr{C}(x^*, \lambda^*)$$

3. Si $x^* \in \mathbb{R}^n$ y λ^* es un multiplicador de Lagrange que satisface las condiciones KKT de modo que

$$w^T \nabla_{xx}^2 \mathscr{L}(x^*, \lambda^*) w > 0, \ w \in \mathscr{L}(x^*, \lambda^*), \ w \neq 0$$

entonces x^* es una solución local estricta para el problema (#).

4. Para $\mu > 0$ cualquiera, el punto $\left(\overline{x}, \overline{\lambda}\right)$ es una solución local del sistema KKT del problema (#) si y solamente si $\nabla \mathscr{L}_A\left(\overline{x}, \overline{\lambda}; \mu\right) = 0$, donde

$$\mathscr{L}_A(x,\lambda;\mu) = f(x) + \sum_{i\in\varepsilon} \lambda_i c_i(x) + \frac{1}{2\mu}\sum_{i\in\varepsilon} c_i^2(x) + \\ + \frac{\mu}{2}\left[\sum_{i\in I}\left(\text{máx}\left\{0, \lambda_i - \frac{c_i(x)}{\mu}\right\}\right)^2 - \lambda_i^2\right]$$

Es decir los puntos KKT del problema (#) corresponden a puntos estacionarios sin restricciones del Lagrangiano Aumentado para cualquier $\mu > 0$.

5. Si x^* es una solución local del problema (#), satisface la condición LICQ y si x^* satisface las condiciones suficientes de segundo orden para un multiplicador de Lagrange λ^*, entonces existe $\overline{\mu}$ tal que para todo $\mu \in \langle 0, \overline{\mu}]$, x^* es un minimizador local estricto de $\mathscr{L}_A(x, \lambda^*, \mu)$ definido por:

$$\mathscr{L}_A(x,\lambda;\mu) = f(x) - \sum_{i\in\varepsilon} \lambda_i c_i(x) + \frac{1}{2\mu}\sum_{i\in\varepsilon} c_i^2(x)$$

6. El problema (#) con restricciones de igualdad y desigualdad podemos transformarlo a un problema con restricciones de igualdad introduciendo variables de holgura t_i de la siguiente forma

$$\min_x f(x) \text{ sujeto a } \begin{cases} c_i(x) = 0 & ,i\in\varepsilon \\ -c_i(x) + t_i^2 = 0 & ,i\in I \end{cases}$$

Resultando que los puntos estacionarios del problema transformado, son también puntos estacionarios del problema de optimización no lineal.

CAPÍTULO VI

RECOMENDACIONES

1. Se recomienda demostrar las condiciones de Karush Kuhn Tucker (KKT) con otras condiciones de calificación que sean más débiles que las condiciones de calificación de independencia lineal (LICQ).

2. Existen problemas especiales de optimización no lineal que no satisfacen las condiciones de calificación usuales y problemas que no satisfacen las condiciones de Karush Kuhn Tucker (KKT), por lo que se recomienda estudiar otros métodos de optimización no lineal especialmente para estos casos.

3. Considerar que existen otros métodos para resolver problemas de optimización no lineal como: PQS (Programación Cuadrática Secuencial), basados en Newton, de filtros, de puntos interiores, de penalizaciones. Todos tienen alguna ventaja sobre otras, pero no existe un método que resuelva todos los problemas (algunos son más rápidos pero ocupan mucha memoria), algunos resuelven más problemas pero son lentos.

4. Existen diferentes algoritmos comerciales basados en los métodos anteriores como LANCELOT y ALGENCAN que son aplicados a problemas de estadística, economía, bioequivalencia, etc. Por lo que se sugiere realizar trabajos de implementación de algoritmos basados en los diferentes métodos de solución.

CAPÍTULO VII

REFERENCIAS

Alves, A. & Wegner, E. (2014). Optimización Continua: Aspectos teóricos y computacionales.*Cengage Learning Ediciones*.

Andreani, R., Ramos, A., Ribeiro, A.A., Secchin, L.D., & Velazco, A.R.(2021). On the convergence of augmented Lagrangian strategies for nonlinear programming. *IMA Journal of Numerical Analysis*, 01-32.doi:https://doi.org/10.1093/imanum/drab021

Beck, A. (2023). Introduction to nonlinear optimization: Theory, algorithms and applica tions with Python and MATLAB (Second edition). *Society for Industrial and Applied Mathematics.*

Bertsekas, D. (1999). Nonlinear Programming (Second edition). *Cambridge: Athena Scientific*. Obtenido de http://world.std.com/ athenasc/index.html

Charmaz, K. (2008). Grounded Theory as an Emergent Method. *Handbook of Emergent Methods*, 155-172.

Croceri, G.M., & Sottosanto, G.N. (2011). Aplicación de un Algoritmo de Lagrangiano Aumentado a la Resolución del Problema de Cuadrados Mínimos No Lineales. *Mecánica Computacional, XXX*, 2023-2036. Obtenido de https://cimec.org.ar/ojs/index.php/mc/article/view/3888

Cui, Y., Ding, C., Li, X., & Zhao, X. (2021). Augmented Lagrangian Methods for Convex Matrix Optimization Problems. *Journal of the Operations Research Society of China*. doi:https://doi.org/10.1007/s40305-021-00346-9

Kanzow, C., Steck, D., & Wachsmuth, D. (2018). An Augmented Lagrangian Method for Optimization.*Society for Industrial and Applied Mathematics*, 56(1), 272-291.

doi:https://doi.org/10.1137/16M1107103

Lages, E. (2004). Análise Real (Volume 2), *Instituto Nacional de Matemática Pura y Aplicada.*

Mehta, R. (2020). Recursive quadratic programming for constrained nonlinear optimization of session throughput in multiple-flow network topologies. *Engineering Reports*, 01-14.doi:https://onlinelibrary.wiley.com/doi/full/10.1002/eng2.12169

Nocedal, J., & Wright, S. (1999). Numerical Optimization. *New York: Springer-Verlag.*

Rojas, F. (2018). Condições de Qualificação, Otimalidade e um Método tipo Lagrangiano Aumentado. *Proceeding Series of the Brazilian Society of Applied and Computational Mathematics*, 06(01), 010339-1-010339-6. doi:10.5540/03.2018.006.01.0339

Soledad, N. (2018). Teoría y Métodos para Problemas de Optimización [Tesis Doctoral, Universidad Nacional de la Plata]. Repositorio Institucional de la UNLP. Obtenido de http://sedici.unlp.edu.ar/handle/10915/70951

Vázquez, M. (2021). Resolución de los problemas de optimización con restricciones mediante los Métodos de Penalización y del Lagrangiano Aumentado.*Revista Digital: Matemática, Educación E Internet*, 21(2), 01-23. doi:https://doi.org/10.18845/rdmei.v21i2.560

Wang, H., Zhang, F., Wang, J., & Rong, Y. (2020). An inexact first-order method for constrained nonlinear.*Optimization Methods and Software*, 01-34. doi:10.1080/10556788.2020.1712601

Yi-Chih, H., Yung-Cheng, L., & Peng-Sheng, Y. (2015). Solving nonlinear constrained optimization problems: An immune evolutionary based two-phase approach. *Applied Mathematical Modelling*, 39(19), 5759-5768. doi:https://doi.org/10.1016/j.apm.2014.12.019

AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Yeni Elizabeth Lauracio Quispe, identificado con DNI 73146460 en mi condición de egresado de:

☒ **Escuela Profesional,** ☐ **Programa de Segunda Especialidad,** ☐ **Programa de Maestría o Doctorado**

Ciencias Físico Matemáticas,

informo que he elaborado el/la ☒ **Tesis** o ☐ **Trabajo de Investigación** denominada:

"MÉTODO LAGRANGIANO AUMENTADO PARA LA SOLUCIÓN DE PROBLEMAS DE OPTIMIZACIÓN NO LINEAL"

para la obtención de ☐**Grado,** ☒ **Título Profesional** o ☐ **Segunda Especialidad.**

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los "Contenidos") que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: https://creativecommons.org/licenses/by-nc-sa/4.0/

En señal de conformidad, suscribo el presente documento.

Puno 23 de julio del 2024

FIRMA (obligatoria)

Huella

DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Yeni Elizabeth Lauracio Quispe,
identificado con DNI 73146460 en mi condición de egresado de:

☒ **Escuela Profesional,** ☐ **Programa de Segunda Especialidad,** ☐ **Programa de Maestría o Doctorado**

Ciencias Físico Matemáticas,

informo que he elaborado el/la ☒ **Tesis** o ☐ **Trabajo de Investigación** denominada:
" MÉTODO LAGRANGIANO AUMENTADO PARA LA SOLUCIÓN DE PROBLEMAS DE OPTIMIZACIÓN NO LINEAL "

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como suyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 23 de julio del 2024

FIRMA (obligatoria)

Huella