

UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE CIENCIAS



TESIS:

**MÉTODO DE PUNTOS INTERIORES PARA OPTIMIZACIÓN CONVEXA NO
DIFERENCIABLE**

PARA OBTENER EL TÍTULO PROFESIONAL DE:
LICENCIADO EN MATEMÁTICA

ELABORADO POR:

JESÚS CERNADES GÓMEZ

ASESOR:

DR. LUIS ERNESTO FLORES LUYO

Lima - Perú

2021

Esta tesis está dedicada a las personas que amo ya que ellas siempre han estado a mi lado en todo momento, a mis pequeños hijos, mi madre Alejandra Gómez Rosales y mi padre Wilbert Valeriano Cernades Maxdeo que ya no está a mi lado.

Índice

1. Introducción	11
2. Nociones preliminares	14
2.1. Definiciones básicas	14
2.2. Dirección de descenso, factible y longitud de paso	18
2.3. Función convexa y consecuencias	21
2.3.1. Propiedades básicas de una función convexa	23
2.3.2. Funciones convexas diferenciables	24
2.3.3. Funciones convexas no diferenciables	25
2.3.4. Optimalidad convexa	28
2.4. Condiciones de Karush-Kuhn-Tucker	29
2.5. Método de Newton para ecuaciones	31
3. Método de Direcciones Factibles de Puntos Interiores para Optimización No Lineal Diferenciable	33
3.1. Método FDIPA	34
3.2. Hipótesis adicionales para la convergencia global	35
3.3. Algoritmo de Puntos Interiores de Direcciones Factibles	39
3.4. Convergencia del algoritmo FDIPA	41
4. Métodos para programación convexas no diferenciables	44
4.1. Método del Subgradiente	45
4.1.1. Deficiencias del método del Subgradiente	50
4.2. Método de Planos Cortantes	52

4.2.1. Ventajas y desventajas del método de Planos Cortantes	54
4.3. Método de Bundle	55
5. Método de Direcciones Factibles para Optimización Convexa no diferenciable	62
5.1. El Método de Direcciones Factibles no Diferenciables	63
5.1.1. Descripción del Método NFDA	63
5.2. Algoritmo de Direcciones Factibles no Diferenciables - NFDA	65
5.3. Análisis de convergencia del algoritmo NFDA	67
5.4. Algunas ideas para solucionar otros tipos de problemas con el método NFDA	79
6. Resultados numéricos	82
6.1. Problemas Test	82
6.2. Resultados obtenidos	95
Bibliografía	98

Lista de figuras

2.1. Los valores de α que satisfacen la desigualdad de Armijo $f(x^k + \alpha d^k) \leq f(x^k) + \sigma \alpha \langle \nabla f(x^k), d^k \rangle$	20
2.2. Los valores de α que satisfacen la desigualdad de Goldstein $\sigma_1 \leq \frac{f(x^k + \alpha d^k) - f(x^k)}{\alpha \langle \nabla f(x^k), d^k \rangle} \leq \sigma_2$	21
2.3. Existencia de un hiperplano soporte sobre un conjunto D	22
2.4. Función convexa	23
3.1. d dirección factible	38
4.1. Condición para escoger la longitud de paso para el método del Sub- gradiente	46
4.2. La dirección $y = -(1, 2)$ que es un anti-subgradiente de f en x pero no es una dirección de descenso de f	51
4.3. Cuatro primeras iteraciones del método de Planos Cortantes	53
4.4. El gráfico muestra un crecimiento de f en la iteración x^{k+1} , pero ella se aleja de la solución óptima	54
4.5. Un paso factible del método de Bundle genera en la iteración x^{k+1} un decrecimiento de f , es decir, $f(x^{k+1}) < f(x^k)$. Además podemos observar del gráfico que x^{k+1} es una buena aproximación a la solución \bar{x} con respecto a la aproximación del método de Planos Cortantes. . .	56

Índice de cuadros

6.1.	it: número de iteraciones factibles, cti : total de pasos nulos, itt : número total de iteraciones, nf: número de veces que la función objetivo y subgradiente fueron calculadas y f^* : valor óptimo encontrado.	95
6.2.	Método NFDA Vs M1FC1 (ϵ-Máximo Descenso), BT (Bundle con Regiones de confianza) y PB (Proximal Bundle). it: número de iteraciones factibles. nf: número de iteraciones que fueron calculados los subgradientes y f . f^* : valor óptimo encontrado.	96

Lista de símbolos

A, B, C	matrices
\mathbb{R}^n	espacio euclideo n -dimensional
$\nabla f(x)$	gradiente de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ en el punto x
$\nabla^2 f(x)$	matriz hessiana de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ en el punto x
$Jf(x)$	matriz jacobiano de $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ en el punto x
$\partial f(x)$	subgradiente de una función convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ en el punto x
$\partial_\epsilon f(x)$	ϵ -subgradiente de la función convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ en el punto x
$\mathcal{N}_D(x)$	el cono normal de un conjunto D convexo no vacío
$\{x_k\}, (x_k) = (x_0, x_1, \dots, x^k, \dots)$	una secuencia de puntos de \mathbb{R}
$\frac{\partial f}{\partial d}(x)$	derivada direccional de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ en el punto x y en la dirección d
$\mathcal{D}_f(x)$	el conjunto de direcciones de descenso de la función f en el punto x
$\mathcal{V}_D(x)$	el conjunto de direcciones viables del conjunto D respecto al punto x
$\bar{B}_\delta(\bar{x})$	la bola cerrada centrada en el punto $\bar{x} \in \mathbb{R}^n$ y radio $\delta > 0$
$L_{f,X}(c)$	el conjunto de nivel de la función f definida en X en c
x_i	la componente i -ésima de $x = (x_1, x_2, \dots, x_n)$

$\text{epi}(f)$	epígrafo de la función f
$\text{int}(\text{epi}(f))$	interior del epígrafo de la función f
$\mathbb{R}^{n \times n}$	espacio de matrices cuadradas en \mathbb{R}
$x^t \cdot y$	producto interno canónico en \mathbb{R}^n
$\ \cdot\ $	la norma euclidiana n-dimensional
\overline{X}	la cerradura del conjunto X
$\text{fr}X$	la frontera del conjunto X
\mathbb{R}_+	el conjunto de los números reales positivos
\mathbb{R}_+^0	el conjunto de los números reales no negativos

Resumen

El objetivo principal de esta tesis es resolver un problema de optimización convexa no diferenciable sin restricciones. La técnica que se aplica para solucionar el problema es un método de puntos interiores realizado por Wilhelm P. Freire, José Herskovits, Mario Tanaka Fo y Alfredo Canelas. Este método consiste en generar una secuencia $\{d^k\}$ de direcciones de descenso y factibles en la región interior del epígrafo de la función objetivo y que a su vez generan una secuencia de puntos $\{x^k\}$ donde cada punto de acumulación es solución del problema primal planteado. Se resolvieron algunos problemas test tomados de la literatura para la ejecución del algoritmo del método presentado y se comprobó la eficacia comparando estos resultados con otros métodos de optimización convexa no diferenciables. Finalmente se terminó dando algunas ideas para la mejora del método estudiado, así como algunas ideas para resolver problemas de optimización no diferenciables con restricciones de igualdad y desigualdad.

Palabras claves: Optimización convexa no diferenciable, Subgradiente, Direcciones de búsqueda.

Abstract

The main objective of this thesis is to solve a non-differentiable convex optimization problem without restrictions. The technique applied to solve the problem is an interior points method carried out by Wilhelm P. Freire, José Herskovits, Mario Tanaka Fo and Alfredo Canelas. This method consists of generating a sequence $\{d^k\}$ of descending and feasible directions in the interior region of the epigraph of the objective function and which in turn generates a sequence of points $\{x^k\}$ where each accumulation point is a solution to the primal problem posed. Some test problems taken from the literature were solved for the execution of the algorithm of the presented method and the efficiency was verified comparing these results with other non-differentiable convex optimization methods. Finally, some ideas for improving the method studied were given, as well as some ideas for solving non-differentiable optimization problems with equality and inequality restrictions.

Keywords: Undifferentiable convex optimization, Subgradient, Search directions.

Capítulo 1

Introducción

Optimización es una línea de la Matemática a gran escala, ya que estudia una variedad de problemas reales que es cada vez más amplia en ciencia e ingeniería. Es por ello que la línea de optimización genera una variedad de métodos y en consecuencia algoritmos para resolver los diferentes tipos de problemas que se puedan originar.

Existen muchos problemas de optimización cuyo interés recae en encontrar puntos donde la función objetivo sea diferenciable o no y ellas se maximicen o minimicen. La optimización de problemas con funciones convexas no necesariamente diferenciables es desarrollado por programación no diferenciable, llamada también optimización no suave, y el cual será el capítulo principal de estudio en esta tesis.

En esta ocasión estudiaremos un método de programación dado en la literatura y el cual fue desarrollado en la década pasada. Se tomó este método por la eficacia que posee en la determinación de una solución a problemas de minimización que se implementaron en esta tesis. Este método ya fue utilizado para resolver problemas de diversos tipos: como en la ingeniería, economía, administración, física, química, vea por ejemplo [6], [7], modelados matemáticamente de la siguiente manera

$$\begin{aligned} \text{mín } & f(x) \\ \text{s.a } & x \in \mathbb{R}^n \end{aligned}$$

Para la resolución de la clase de problemas de Programación Convexa no Diferenciables es necesario el uso de nuevas técnicas que reemplazan los métodos clásicos de estudio en

el cálculo diferenciable. Tales técnicas provienen de una de las áreas de la matemática llamada Análisis Convexo.

Los problemas que abarcaremos en la tesis son aquellos donde la función objetivo es convexa y no necesariamente diferenciable. Es por ello que desarrollaremos el método llamado NFDA (A Feasible Directions Method for Nonsmooth Convex Optimization) y trataremos de dar algunas recomendaciones para mejorar dicho método. Este método resuelve problemas que minimizan funciones convexas no diferenciables sin restricción. La técnica que emplea el método es basada en el método de Planos Cortantes (vea [30] [29]) el cual consiste en hacer una aproximación lineal a la función objetivo generando un problema aproximado del problema original. Para determinar el valor óptimo de la función objetivo se usa un método de puntos interiores al epígrafo de la función objetivo aproximada con direcciones factibles y de descenso, que es un método para problemas de optimización no lineal diferenciable el cual evita resolver un problema basado en métodos de penalidades o barreras, éste método genera un algoritmo que es fácil de implementar y eficiente con una convergencia superlineal. El método de puntos interiores en que se basa el método NFDA es llamado FDIPA (A Feasible Directions Interior Point Technique For Nonlinear Optimization) que fue desarrollado por Herskovits en [16]. Este método consiste que en cada iteración una dirección de descenso y factible es obtenida resolviendo apenas dos sistemas lineales con la misma matriz, el cual es no singular en cada iteración del algoritmo.

El método NFDA, debido a que usa aproximaciones lineales tanto de la función objetivo como de las restricciones, trae dificultades computacionales. Es por ello que este algoritmo es llamado un método de memoria (métodos de Subgradiente), y por ello que puede ser mejorado usando algunas técnicas basadas por los métodos de Bundle, el cual veremos en el capítulo 4. Finalmente, podemos concluir diciendo que este método puede resolver problemas de optimización convexa con restricciones de desigualdad convexa como también problemas no convexos.

El algoritmo fue programado en Octave 6.2.0 y testeado utilizando un conjunto de problemas test encontrados en la literatura [27]. Los resultados obtenidos fueron satisfactorios y los subgradientes tomados en cada iteración fueron escogidos de manera

apropiada aunque aún se busca cuál de todos ellos puede ser escogida para mejorar la optimalidad en la obtención de la solución del problema.

Este trabajo está organizado en seis capítulos, siendo este el primero. En el capítulo 2 se dará algunas definiciones y resultados importantes que serán de ayuda para nuestro objetivo y utilizados en los demás capítulos. En el capítulo 3 introduciremos el problema de optimización no lineal diferenciable con restricciones de desigualdad, que será resuelto por el método FDIPA. En el capítulo 4 haremos una rápida introducción a los métodos de optimización convexa no diferenciable y veremos algunas dificultades que se presentan por la falta de diferenciabilidad. En el capítulo 5 se estudiará el método NFDA propuesto por Freire, ver [30], que es el estudio de la tesis. En este capítulo se estudiará de qué manera se puede hacer variar el método para resolver problemas no convexos y con restricciones de desigualdad. En el capítulo 6 comprobaremos la eficacia del algoritmo generado por el método NFDA. Los problemas test son muy utilizados para la comprobación de la eficacia y rapidez al obtener la solución óptima de un problema de optimización. Para finalizar daremos algunas conclusiones del método y como podría ser mejorado, y finalmente presentamos la bibliografía usada.

Capítulo 2

Nociones preliminares

En este capítulo presentamos algunas definiciones básicas y propiedades que serán de gran utilidad para el desarrollo de la tesis. Estos resultados encontrados por ejemplo [1], [2], [4], [21] y [28], serán esenciales para el estudio de los métodos de optimización diferenciable y no diferenciable los cuales veremos en los capítulos posteriores. Un resultado muy importante que se realizará con detalle en este capítulo será la condición necesaria y suficiente de optimalidad para un problema de optimización convexa no diferenciable y que será de gran ayuda el demostración de la convergencia del algoritmo que desarrollaremos en el capítulo 5.

2.1. Definiciones básicas

Empezaremos con la definición del problema primal, conceptos de optimalidad para problemas diferenciables (cuyos algoritmos poseen criterios de parada). Seguidamente veremos conceptos de optimización no diferenciable los cuales requieren de otro estudio en relación a los métodos diferenciables, en especial por su criterio de parada. Es por ello que los conceptos de subgradiente y subdiferencial son importantes en esta tesis. Estos conceptos sustituyen algunos resultados de diferenciabilidad y conllevan a nuevas propiedades de las funciones convexas no diferenciables. Por tanto los algoritmos de optimización no diferenciable requieren nuevos criterios de paradas, diferentes a los problemas de optimización diferenciable.

El Problema Primal

Dado una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, y $D \subseteq \mathbb{R}^n$ un subconjunto, definimos el problema de optimización, llamado problema primal como

$$(P) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s. a} & x \in D, \end{cases}$$

donde D es llamado el conjunto factible del problema (P) , los elementos de D son llamados puntos factibles y f es llamada función objetivo. En esta tesis usaremos la condición de convexidad para la función f y el conjunto factible $D = \mathbb{R}^n$.

El problema primal (P) , es un problema de optimización convexa, si D es convexo y f es una función convexa sobre D .

Definición 2.1. Se dice que un punto $\bar{x} \in D \subset \mathbb{R}^n$ es:

1. *Mínimo local para (P) , si existe $\delta > 0$ y una bola abierta $B_\delta(\bar{x})$ tal que*

$$f(\bar{x}) \leq f(x), \quad \forall x \in D \cap B_\delta(\bar{x}).$$

2. *Mínimo global para (P) , si*

$$f(\bar{x}) \leq f(x), \quad \forall x \in D.$$

Definición 2.2. Diremos que $\bar{v} \in [-\infty, +\infty)$ es el valor óptimo del problema (P) , si

$$\bar{v} = \inf_{x \in D} f(x),$$

En muchas aplicaciones a finanzas, economía, etc, vea por ejemplo [20], [28] y [39], el conjunto factible $D \subset \mathbb{R}^n$ es descrito por sistemas de desigualdades lineales, no lineales sobre un conjunto X de \mathbb{R}^n , es decir,

$$D = \begin{cases} g_i(x) \leq 0 & \text{para } i = 1, \dots, m \\ h_j(x) = 0 & \text{para } j = 1, \dots, l \\ x \in X \end{cases}$$

donde las funciones $g_1, \dots, g_m, h_1, \dots, h_l$ están definidas en \mathbb{R}^n y X puede ser acotado, compacto, convexo, etc.

Definición 2.3. Sean $D \subseteq \mathbb{R}^n$ un subconjunto, $f : D \rightarrow \mathbb{R}$ y $x_0 \in D$. Diremos que f es continua en $x_0 \in D$ si $\forall \epsilon > 0, \exists \delta > 0$ tal que $\forall x \in D, \|x - x_0\| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon$.

Teorema 2.1 (Teorema de Weierstrass, vea [1]). Sea $D \subseteq \mathbb{R}^n$ un subconjunto compacto no vacío y sea $f : D \rightarrow \mathbb{R}$ una función continua. Entonces, f alcanza su mínimo y su máximo valor en el conjunto D y el problema primal (P) tiene solución global.

Definición 2.4. Sea $D \subseteq \mathbb{R}^n$ un subconjunto. El conjunto de nivel de una función $f : D \rightarrow \mathbb{R}$ asociada a $c \in \mathbb{R}$, es aquel conjunto de puntos dado por

$$L_{f,D}(c) = \{x \in D \mid f(x) \leq c\}.$$

Corolario 2.1.1. Sea $D \subseteq \mathbb{R}^n$ un subconjunto no vacío y sea $f : D \rightarrow \mathbb{R}$ continua sobre D . Suponiendo que existe $c \in \mathbb{R}$ tal que el conjunto de nivel $L_{f,D}(c)$ sea no vacío y compacto, el problema primal (P) posee un mínimo global.

Definición 2.5. Sea $D \subseteq \mathbb{R}^n$ un subconjunto y sea $f : D \rightarrow \mathbb{R}$. Se dice que f es Lipschitziana cuando existe una constante $k > 0$ tal que para cualquier par de puntos $x, y \in D$, se tiene que $|f(x) - f(y)| \leq k\|x - y\|$.

Es claro que f es continua si es Lipschitziana.

Definición 2.6. Sean $f : D \rightarrow \mathbb{R}$ una función definida en un conjunto abierto $D \subseteq \mathbb{R}^n$, $a \in D$ y $v \in \mathbb{R}^n$. Se define la derivada direccional de f en a y en la dirección v al límite (si existe),

$$\frac{\partial f}{\partial v}(a) = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t}.$$

En tal caso diremos f es derivable en el punto a y en la dirección v .

Cuando el vector $v = e_i = (0, 0, \dots, \underbrace{1}_{i\text{-ésimo término}}, \dots, 0)$, para $i = 1, \dots, n$, la derivada direccional

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \rightarrow 0} \frac{f(a + te_i) - f(a)}{t}$$

es llamada derivada parcial (con respecto a la i -ésima variable).

Definimos la gradiente de una función f en el punto a como el vector $\nabla f(a) = \left[\frac{\partial f(a)}{\partial x_1}, \dots, \frac{\partial f(a)}{\partial x_n} \right]^t$, siempre que existen las derivadas parciales.

Definición 2.7. Sea $f : D \rightarrow \mathbb{R}$ una función, con $D \subseteq \mathbb{R}^n$ un subconjunto abierto no vacío, y sea $a \in D$. Se dice que f es diferenciable en el punto a si existen escalares reales A_1, A_2, \dots, A_n (que dependen de a) tales que para todo $v = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$, con $a + v \in D$, se cumple

$$f(a + v) = f(a) + \sum_{i=1}^n A_i \alpha_i + r(v),$$

donde r es una función real definida en $D - \{a\} = \{v \mid a + v \in D\}$ tal que $\lim_{v \rightarrow 0} \frac{r(v)}{\|v\|} = 0$.

Es fácil comprobar que la existencia de las constantes reales A_1, A_2, \dots, A_n son en realidad las derivadas parciales de f en el punto a .

Diremos que f es diferenciable en D , si f es diferenciable en cada punto de D , vea algunos resultados de diferenciability por ejemplo en [21].

La función $f : D \rightarrow \mathbb{R}$, es dos veces diferenciable en $a \in D$, si $\frac{\partial f}{\partial x_i}$ son diferenciables en a , para $i = 1, \dots, n$ y $D \subseteq \mathbb{R}^n$ abierto.

Si la función f es dos veces diferenciable en $a \in D$, entonces la matriz Hessiana de f en a , definida por

$$\nabla^2 f(a) = \begin{pmatrix} \frac{\partial^2 f(a)}{\partial x_1^2} & \frac{\partial^2 f(a)}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(a)}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(a)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(a)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(a)}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(a)}{\partial x_1 \partial x_n} & \frac{\partial^2 f(a)}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(a)}{\partial x_n^2} \end{pmatrix}_{n \times n}.$$

es simétrica.

Sea $f = (f_1, \dots, f_m) : D \rightarrow \mathbb{R}^m$ una función, si f_i es diferenciable en el punto $a \in D$, para $i = 1, \dots, m$ y $D \subseteq \mathbb{R}^n$ abierto, definimos la matriz Jacobiana por $Jf(a) = \left(\frac{\partial f_i}{\partial x_j}(a) \right)_{m \times n}$. Si la función $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, con D abierto se puede escribir de

la forma $f(x) = \psi(x, \bar{z})$, donde $\psi : \mathbb{R}^n \times Z \rightarrow \mathbb{R}$ y Z un subconjunto cualquiera, denotamos $\nabla f(a) = \psi'_x(a, \bar{z})$, con \bar{z} fijo.

Definición 2.8. Sean $D \subseteq \mathbb{R}^n$ un subconjunto no vacío y $f : D \rightarrow \mathbb{R}$. Sea $\bar{x} \in D$ y d un vector no nulo de \mathbb{R}^n tal que $\bar{x} + \lambda d \in D$ para $\lambda > 0$ y suficientemente pequeño. La derivada de f en el punto \bar{x} y la dirección d , denotado por $f'(\bar{x}; d)$, es dado por el siguiente límite si existe:

$$f'(\bar{x}; d) = \lim_{\lambda \rightarrow 0^+} \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda}.$$

En la siguiente sección definiremos y mencionaremos ciertos resultados importantes, para el desarrollo del método de puntos interiores que desarrollaremos en el siguiente capítulo: dirección de descenso, factible y longitud de paso, que pueden ser encontrados por ejemplo [1], [4] y [28].

2.2. Dirección de descenso, factible y longitud de paso

Definición 2.9. Un conjunto $D \subseteq \mathbb{R}^n$ es llamado conjunto convexo si para cualquier $x, y \in D$ y para todo $\lambda \in [0, 1]$, se tiene $\lambda x + (1 - \lambda)y \in D$.

Definición 2.10. Un conjunto $D \subseteq \mathbb{R}^n$ es llamado cono cuando

$$d \in D \Rightarrow td \in D, \quad \forall t \in \mathbb{R}_+.$$

Definición 2.11. Sean $D \subseteq \mathbb{R}^n$ un conjunto convexo y $\bar{x} \in D$. El cono normal (cono de direcciones normales) en el punto \bar{x} en relación al conjunto D es el conjunto

$$\mathcal{N}_D(\bar{x}) = \{d \in \mathbb{R}^n \mid \langle d, x - \bar{x} \rangle \leq 0 \quad \forall x \in D\}.$$

Definición 2.12. Diremos que $d \in \mathbb{R}^n$ es una dirección factible en relación al conjunto D en el punto $\bar{x} \in D$, si existe $\epsilon > 0$ tal que

$$\bar{x} + td \in D, \quad \forall t \in [0, \epsilon]$$

En adelante denotaremos por $\mathcal{V}_D(\bar{x})$ al conjunto de todas las direcciones factibles sobre el conjunto D en el punto $\bar{x} \in D$.

Definición 2.13. La cápsula convexa de un conjunto $D \subset \mathbb{R}^n$, denotada por $\text{conv}(D)$, es el menor conjunto convexo en \mathbb{R}^n que contiene a D .

Definición 2.14. Diremos que $d \in \mathbb{R}^n$ es una dirección de descenso de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ en el punto $\bar{x} \in \mathbb{R}^n$, si existe $\epsilon > 0$ tal que

$$f(\bar{x} + td) < f(\bar{x}), \quad \forall t \in \langle 0, \epsilon \rangle.$$

Denotaremos por $\mathcal{D}_f(\bar{x})$ al conjunto de todas las direcciones de descenso de f en el punto \bar{x} .

El siguiente lema nos da un criterio para identificar direcciones de descenso.

Lema 2.1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable en el punto $\bar{x} \in \mathbb{R}^n$. Entonces

1. Para todo $d \in \mathcal{D}_f(\bar{x})$, se tiene $\langle \nabla f(\bar{x}), d \rangle \leq 0$.
2. Si $d \in \mathbb{R}^n$ satisface $\langle \nabla f(\bar{x}), d \rangle < 0$, entonces $d \in \mathcal{D}_f(\bar{x})$.

Sea d una dirección de descenso, entonces existe $\epsilon > 0$ tal que $f(x + td) < f(x)$ para todo $t \in \langle 0, \epsilon \rangle$ por lo tanto debemos encontrar algún t de manera que reduzca a la función objetivo en el punto $x + td$. La existencia de este valor escalar t , el cual es llamado longitud de paso, conlleva a ver la forma de cómo obtenerlo y de manera que la función decrezca de manera aceptable. Para dicha finalidad existen reglas de Armijo, Goldstein u otros, que las presentamos a continuación.

La regla de Armijo

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable en el punto x^k y d^k un vector de \mathbb{R}^n tal que $\langle \nabla f(x^k), d^k \rangle < 0$. Fijamos los parámetros σ y θ en $\langle 0, 1 \rangle$, y $\hat{\alpha} > 0$. La regla de Armijo consiste en calcular la longitud de paso de manera aproximada en la que la función decrezca suficientemente en la iteración $k + 1$ con respecto a la iteración k , es decir:

Sea $\alpha = \hat{\alpha}$.

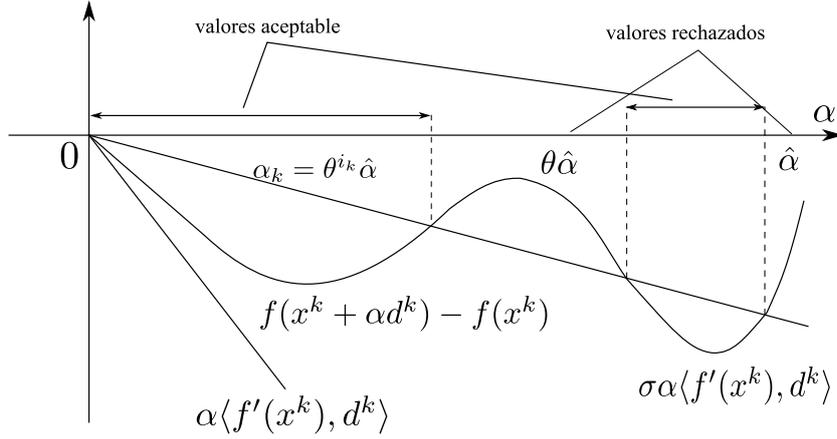


Figura 2.1: **Los valores de α que satisfacen la desigualdad de Armijo $f(x^k + \alpha d^k) \leq f(x^k) + \sigma \alpha \langle \nabla f(x^k), d^k \rangle$**

1. Verificamos si la desigualdad

$$f(x^k + \alpha d^k) \leq f(x^k) + \sigma \alpha \langle \nabla f(x^k), d^k \rangle \quad (2.1)$$

se satisface o no.

2. Si (2.1) no se satisface, tomamos $\alpha = \theta \alpha$ y retornamos al paso 1. En caso contrario, aceptamos $\alpha_k = \alpha$ como el valor de la longitud de paso.

Para ver con más detalles la regla de Armijo y encontrar alguna cota inferior de la longitud de paso, vea por ejemplo [1].

La regla de Goldstein

Para el caso de la regla de Goldstein se tienen las mismas condiciones para la función f , y para calcular la longitud de paso, la regla de Goldstein debe satisfacer la siguiente desigualdad

$$\sigma_1 \leq \frac{f(x^k + \alpha d^k) - f(x^k)}{\alpha \langle \nabla f(x^k), d^k \rangle} \leq \sigma_2, \quad (2.2)$$

donde $0 < \sigma_1 < \sigma_2 < 1$ son parámetros dados. Aquí la primera desigualdad es de Armijo, con $\sigma = \sigma_1$.

Existen otras reglas inexactas y exactas para encontrar la longitud de paso, ver por ejemplo [1] y [28].

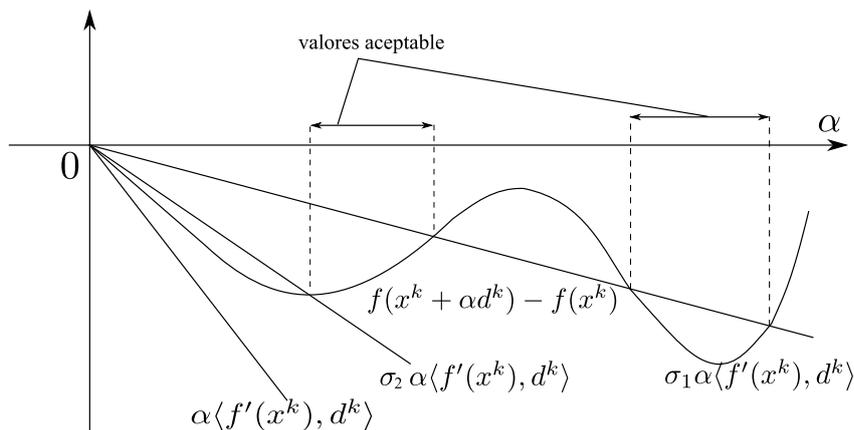


Figura 2.2: **Los valores de α que satisfacen la desigualdad de Goldstein**

$$\sigma_1 \leq \frac{f(x^k + \alpha d^k) - f(x^k)}{\alpha \langle \nabla f(x^k), d^k \rangle} \leq \sigma_2$$

Definición 2.15. Un hiperplano H en \mathbb{R}^n es una colección de puntos de la forma

$$H(p, \alpha) = \{x \in \mathbb{R}^n \mid p^t x = \alpha\}$$

donde p es un vector no nulo y α un escalar real.

De la definición 2.15, tenemos que un hiperplano H define dos semiespacios cerrados $H^+(p, \alpha) = \{x \in \mathbb{R}^n \mid p^t x \geq \alpha\}$ y $H^-(p, \alpha) = \{x \in \mathbb{R}^n \mid p^t x \leq \alpha\}$ o dos semiespacios abiertos $\{x \in \mathbb{R}^n \mid p^t x > \alpha\}$ y $\{x \in \mathbb{R}^n \mid p^t x < \alpha\}$.

Definición 2.16. Sea D un subconjunto no vacío de \mathbb{R}^n , y sea $\bar{x} \in \text{fr}D$. Un hiperplano $H(p, \alpha) = \{x \in \mathbb{R}^n \mid p^t x = \alpha\}$, donde $\alpha = p^t \bar{x}$, es llamado un hiperplano soporte del subconjunto D sobre \bar{x} si $D \subset H^+$, esto es, $p^t x \geq \alpha$ para cada $x \in D$ o $D \subset H^-$, es decir, $p^t x \leq \alpha$ para cada $x \in D$.

Teorema 2.2. Sea D un subconjunto convexo no vacío de \mathbb{R}^n , y sea $\bar{x} \in \text{fr}D$. Existe un hiperplano soporte H en \bar{x} , esto es, existe un vector no nulo p tal que $p^t x \leq \alpha$, donde $\alpha = p^t \bar{x}$, para cada $x \in \bar{D}$ (Ver [4], Teorema 2.4.7).

2.3. Función convexa y consecuencias

En esta sección veremos algunas propiedades de las funciones convexas diferenciables y no diferenciables, así como la condición necesaria y suficiente de optimalidad

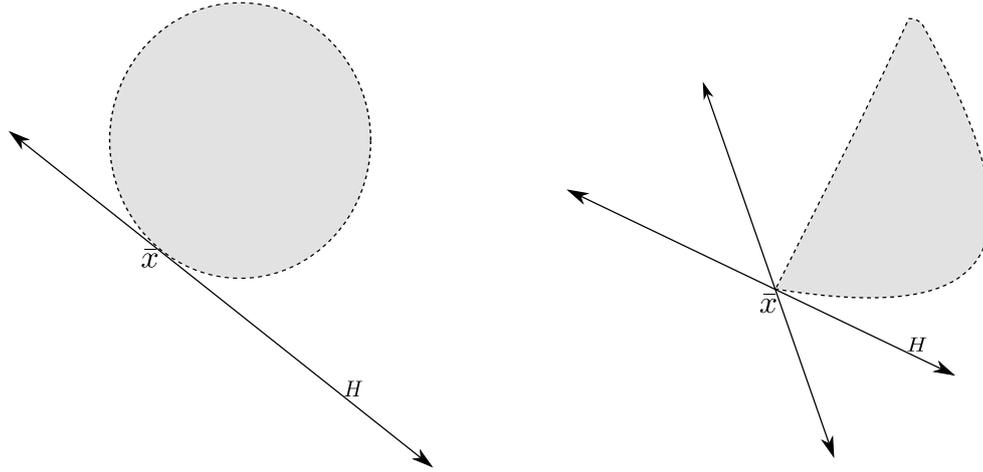


Figura 2.3: **Existencia de un hiperplano soporte sobre un conjunto D**

para un problema de optimización convexa.

Definición 2.17. Sean $D \subseteq \mathbb{R}^n$ es un subconjunto convexo no vacío y $f : D \rightarrow \mathbb{R}$. La función f es llamada:

1. convexa en D si

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

para cada $x_1, x_2 \in D$ y para cualquier $\lambda \in [0, 1]$.

2. estrictamente convexa en D si

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2),$$

para cualquier $x_1 \neq x_2 \in D$ y para cualquier $\lambda \in \langle 0, 1 \rangle$.

3. fuertemente convexa en D con módulo $\lambda > 0$, cuando para cualquier $x, y \in D$ y cualquier $\lambda \in [0, 1]$, se tiene

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \alpha)f(y) - \gamma\lambda(1 - \lambda)\|x - y\|^2.$$

4. cóncava sobre D si la función $-f$ es convexa sobre D .

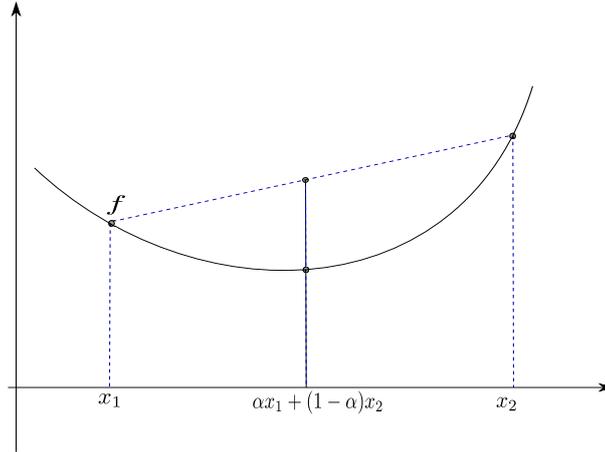


Figura 2.4: **Función convexa**

Definición 2.18. Sea $D \subseteq \mathbb{R}^n$ un subconjunto convexo. Se dice que la función vectorial de variable vectorial $g = (g_1, g_2, \dots, g_m) : D \rightarrow \mathbb{R}^m$ es convexa en D si todas las funciones $g_i : D \rightarrow \mathbb{R}$, $i = 1, \dots, m$, son convexas sobre D .

Definición 2.19. Sean $D \subseteq \mathbb{R}^n$ un subconjunto y $f : D \rightarrow \mathbb{R}$. El epígrafo de f , denotado por $\text{epi}(f)$, es un subconjunto de $\mathbb{R}^n \times \mathbb{R}$

$$\text{epi}(f) = \{(x, y) \in D \times \mathbb{R} \mid f(x) \leq y\}.$$

2.3.1. Propiedades básicas de una función convexa

En esta parte desarrollaremos algunas propiedades de una función f convexa, estrictamente convexa y fuertemente convexa, que serán de mucha importancia en el desarrollo del capítulo 5, cuyos resultados pueden ser encontradas por ejemplo en [1], [4].

Proposición 2.1. Sean $D \subseteq \mathbb{R}^n$ un subconjunto convexo, $f : D \rightarrow \mathbb{R}$ convexa y un escalar $\alpha \in \mathbb{R}$. Entonces el conjunto de nivel

$$L_{f,D}(\alpha) = \{x \in D \mid f(x) \leq \alpha\},$$

es un conjunto convexo.

Teorema 2.3. Sea $D \subseteq \mathbb{R}^n$ un subconjunto convexo no vacío, y $f : D \rightarrow \mathbb{R}$. Entonces f es convexa en D si y solo si $\text{epi}(f)$ es un subconjunto convexo de $\mathbb{R}^n \times \mathbb{R}$.

Proposición 2.2 (Convexidad del supremo de funciones convexas). Sean $D \subseteq \mathbb{R}^n$ un subconjunto convexo y $f_i : D \rightarrow \mathbb{R}$, $\forall i \in I$, funciones convexas en D , donde I es un conjunto de índices. Supongamos que exista $\beta \in \mathbb{R}$ tal que $f_i(x) \leq \beta$, $\forall x \in D$ e $i \in I$.

Entonces la función $f : D \rightarrow \mathbb{R}$ definida por $f(x) = \sup_{i \in I} f_i(x)$, $\forall x \in D$, es convexa.

Proposición 2.3. Sea $\Omega \subseteq \mathbb{R}^n$ un subconjunto convexo, y sean $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ una función convexa y $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$ una función afín. Entonces el conjunto

$$D = \{x \in \Omega \mid h(x) = 0, g(x) \leq 0\}$$

es convexo.

Teorema 2.4 (Continuidad de funciones convexas). Sea $D \subseteq \mathbb{R}^n$ un subconjunto convexo, abierto y no vacío, y sea $f : D \rightarrow \mathbb{R}$ una función convexa. Entonces f es continua en D .

Teorema 2.5. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa, y supongamos que exista $c \in \mathbb{R}$ de manera que el conjunto de nivel

$$L_{f, \mathbb{R}^n}(c) = \{x \in \mathbb{R}^n \mid f(x) \leq c\}$$

sea no vacío y acotado. Entonces $L_{f, \mathbb{R}^n}(t)$ es acotado para cualquier $t \in \mathbb{R}$.

Teorema 2.6 (Compacidad de conjuntos de nivel de una función fuertemente convexa). Supongamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sea fuertemente convexa sobre \mathbb{R}^n . El conjunto de nivel

$$L_{f, \mathbb{R}^n}(c) = \{x \in \mathbb{R}^n \mid f(x) \leq c\}$$

es compacto para cualquier $c \in \mathbb{R}$.

2.3.2. Funciones convexas diferenciables

A continuación desarrollaremos algunas propiedades importantes acerca de las funciones convexas diferenciables, cuyos resultados pueden ser encontrados por ejemplo en [2] y [17].

Teorema 2.7. Sean $D \subseteq \mathbb{R}^n$ un subconjunto abierto y convexo y $f : D \rightarrow \mathbb{R}$ diferenciable sobre D . Los siguientes resultados son equivalentes:

1. f es convexa.
2. $f(y) \geq f(x) + \nabla f(x)^t \cdot (y - x)$, para todo $x, y \in D$.
3. $\nabla f(x)^t \cdot (y - x) \leq \nabla f(y)^t \cdot (y - x)$, para todo $x, y \in D$.

Si además f es dos veces diferenciable sobre D , las propiedades anteriores también son equivalentes

4. $\langle \nabla^2 f(x)d, d \rangle \geq 0 \quad \forall x \in D, \forall d \in \mathbb{R}^n$.

2.3.3. Funciones convexas no diferenciables

En esta sección nos enfocaremos en el desarrollo de algunas definiciones y propiedades básicas del subgradiente de una función f que sustituye el concepto de derivada, para problemas de programación no diferenciable de funciones convexas.

Teorema 2.8 (Derivada de una función convexa). Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa. Entonces para todo $\bar{x} \in \mathbb{R}^n$, existen las derivadas en cualquier dirección $d \in \mathbb{R}^n$. Además,

$$f(\bar{x} + \alpha d) \geq f(\bar{x}) + \alpha f'(\bar{x}; d), \quad \forall \alpha \in \mathbb{R}_+.$$

Proposición 2.4. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa. Entonces para cualquier secuencia de puntos $\{x^k\} \rightarrow x$, y direcciones $\{d^k\} \rightarrow d$ ($k \rightarrow \infty$), tenemos que

$$\limsup_{k \rightarrow \infty} f'(x^k; d^k) \leq f'(x; d).$$

Además, si f es diferenciable en \mathbb{R}^n , ∇f es continua en \mathbb{R}^n .

Teorema 2.9. Sea $D \subseteq \mathbb{R}^n$ un subconjunto convexo no vacío, y sea $f : D \rightarrow \mathbb{R}$ una función convexa. Entonces para cada $\bar{x} \in \text{int}(D)$, existe un vector s tal que el hiperplano

$$H = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} \mid y = f(\bar{x}) + s^t \cdot (x - \bar{x})\}$$

soporta al $\text{epi}(f)$ en $(\bar{x}, f(\bar{x}))$. En particular $f(x) \geq f(\bar{x}) + s^t \cdot (x - \bar{x})$ para todo $x \in D$,

El teorema anterior demuestra la existencia de un vector s , llamado subgradiente de la función convexa f en el punto $\bar{x} \in \text{int}(D)$. Esto motiva a la siguiente definición.

Definición 2.20. Sea $D \subseteq \mathbb{R}^n$ un subconjunto convexo no vacío y sea $f : D \rightarrow \mathbb{R}$ una función convexa. Entonces s es llamado subgradiente de f en el punto $\bar{x} \in \text{int}(D)$, si

$$f(x) \geq f(\bar{x}) + s^t \cdot (x - \bar{x}), \quad \forall x \in D.$$

El conjunto de todas las subgradientes de una función f en \bar{x} , denotado por $\partial f(\bar{x})$, es llamado el subdiferencial de f en \bar{x} , es decir,

$$\partial f(\bar{x}) = \{s \in \mathbb{R}^n \mid f(x) \geq f(\bar{x}) + s^t \cdot (x - \bar{x}), \forall x \in D\}$$

A continuación presentamos algunos importantes resultados de subdiferencial que serán tomados para el desarrollo de los capítulos 4 y 5, y cuyas demostraciones pueden ser encontradas por ejemplo en [1], [4] y [28].

Proposición 2.5. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa y sea $x \in \mathbb{R}^n$, entonces

$$s \in \partial f(x) \Leftrightarrow \left\{ s \in \mathbb{R}^n \mid f'(x; d) \geq s^t \cdot d, \quad \forall d \in \mathbb{R}^n \right\}.$$

Teorema 2.10 (El valor óptimo sobre el subdiferencial). Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa. Entonces para todo $\bar{x} \in \mathbb{R}^n$, el conjunto $\partial f(\bar{x})$ es convexo, compacto y no vacío. Además, para todo $d \in \mathbb{R}^n$ se tiene $f'(\bar{x}; d) = \max_{y \in \partial f(\bar{x})} \langle y, d \rangle$.

Proposición 2.6. Una función convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$, es diferenciable en el punto $\bar{x} \in \mathbb{R}^n$ si y solo si el conjunto $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.

La siguiente proposición muestra el comportamiento de continuidad de la subdiferencial de una función f convexa.

Proposición 2.7. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa, una secuencia $\{x^k\} \rightarrow \bar{x}$ ($k \rightarrow \infty$) y $y^k \in \partial f(x^k)$ para todo k . Entonces la secuencia $\{y^k\}$ es acotada y todos sus puntos de acumulación pertenecen a $\partial f(\bar{x})$.

A continuación mencionamos un resultado importante del análisis convexo el cual muestra la manera de como podemos escoger un subgradiente en un punto x de una función convexa. Este resultado que es de mucho interés para la implementación del algoritmo que se desarrollará en el capítulo 5.

Teorema 2.11. *Supongamos que Z es un conjunto compacto cualquiera y que $\psi : \mathbb{R}^n \times Z \rightarrow \mathbb{R}$ es una función continua tal que $\psi(\cdot, z) : \mathbb{R}^n \rightarrow \mathbb{R}$ sea convexa para todo $z \in Z$. Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$, definida por $f(x) = \max_{z \in Z} \psi(x, z)$, entonces:*

a) *La función f es convexa en \mathbb{R}^n , y para todo $x \in \mathbb{R}^n$, tenemos*

$$f'(x; d) = \max_{\bar{z} \in Z(x)} \psi'(x, \bar{z}; d) \quad \forall d \in \mathbb{R}^n,$$

donde $Z(x) = \{\bar{z} \in Z \mid f(x) = \psi(x, \bar{z}) = \max_{z \in Z} \psi(x, z)\}$. En particular, si $Z(x) = \{\bar{z}\}$ y $\psi(\cdot, \bar{z})$ es diferenciable en el punto $x \in \mathbb{R}^n$, entonces f es diferenciable en x y $\nabla f(x) = \nabla \psi_x(x, \bar{z})$.

b) *Si $\psi(\cdot, z)$ es diferenciable para todo $z \in Z$ y $\nabla \psi_x(x, \cdot)$ es continua en Z , para todo $x \in \mathbb{R}^n$, entonces $\partial f(x) = \text{conv}\{\nabla \psi_x(x, \bar{z}) \mid \bar{z} \in Z(x)\}$, $x \in \mathbb{R}^n$.*

Del teorema anterior tenemos, si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa definida por $f(x) = \max_{i=1, \dots, p} f_i(x)$, donde $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ son funciones convexas diferenciables para $i = 1, \dots, p$, y definiendo el conjunto de índices activos para $x \in \mathbb{R}^n$

$$I(x) = \{i = 1, \dots, p \mid f(x) = f_i(x)\},$$

entonces

$$\partial f(x) = \text{conv}\{\nabla f_i(x), i \in I(x)\} = \left\{ y \in \mathbb{R}^n \mid \begin{array}{l} y = \sum_{i \in I(x)} \alpha_i \nabla f_i(x), \quad \sum_{i \in I(x)} \alpha_i = 1, \\ \alpha_i \in \mathbb{R}_+, i \in I(x) \end{array} \right\}$$

Definición 2.21. *Sean $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa y $\epsilon \in \mathbb{R}_+$. Decimos que $s \in \mathbb{R}^n$ es un ϵ -subgradiente de f en el punto $\bar{x} \in \mathbb{R}^n$ cuando $f(x) \geq f(\bar{x}) + s^t \cdot (x - \bar{x}) - \epsilon$, $\forall x \in \mathbb{R}^n$.*

El conjunto de todos los ϵ -subgradiientes de f en \bar{x} , denotado por $\partial_\epsilon f(\bar{x})$, es llamado el ϵ -subdiferencial de f en \bar{x} . Si $\epsilon = 0$ es llamado simplemente subdiferencial.

Teorema 2.12 (Compacidad del ϵ -subdiferencial). *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa. Entonces para todo $x \in \mathbb{R}^n$ y todo $\epsilon \in \mathbb{R}_+$, el conjunto $\partial_\epsilon f(x)$ es convexo, compacto y no vacío.*

2.3.4. Optimalidad convexa

En esta sección desarrollamos el teorema de condición necesaria y suficiente de optimalidad para funciones convexas diferenciables y no diferenciables. Este último es de importancia para el estudio de la convergencia global del método que desarrollaremos en el capítulo 5. Las demostraciones de estos resultados pueden ser encontradas por ejemplo en [1] y [28].

Teorema 2.13. *En un problema de optimización convexa (P) , todo minimizador local es global. Si f es estrictamente convexa, entonces no puede existir más de un minimizador.*

Corolario 2.3.1. *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función fuertemente convexa y $\emptyset \neq D \subseteq \mathbb{R}^n$ un subconjunto convexo y cerrado. Entonces la función alcanza su mínimo en el conjunto D y además es único.*

Teorema 2.14. *Sean $D \subset \mathbb{R}^n$ un conjunto convexo y $f : \Omega \rightarrow \mathbb{R}^n$ una función convexa y diferenciable sobre el conjunto abierto $\Omega \subseteq \mathbb{R}^n$. Entonces $\bar{x} \in D$ es una solución óptima de (P) si y solamente si*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \forall x \in D,$$

o equivalentemente,

$$-\nabla f(\bar{x}) \in \mathcal{N}_D(\bar{x}).$$

Además, si D es abierto, \bar{x} es una solución óptima del problema (P) si y solamente si $\nabla f(\bar{x}) = 0$.

El siguiente teorema garantiza la condición necesaria y suficiente de optimalidad

local del problema (P) de optimización convexa no diferenciable, y que es de suma importancia en la convergencia del algoritmo que se desarrollará en el capítulo 5.

Teorema 2.15. *Sean $D \subseteq \mathbb{R}^n$ un subconjunto convexo y $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa. Entonces f alcanza su valor mínimo en el punto $\bar{x} \in D$ si y solo si*

$$\exists y \in \partial f(\bar{x}) \text{ tal que } 0 \in \partial f(\bar{x}) + \mathcal{N}_D(\bar{x}).$$

En el caso que D es todo \mathbb{R}^n se tiene que $0 \in \partial f(\bar{x})$.

En el caso que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa definida por $f(x) = \max_{i=1, \dots, p} f_i(x)$, donde $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ son funciones convexas diferenciables sobre \mathbb{R}^n , para $i = 1, \dots, p$, tenemos del teorema 2.11 y 2.15 que \bar{x} es un minimizador de f sobre \mathbb{R}^n si y solo si existen $\alpha_i \in \mathbb{R}_+$, $i \in I(\bar{x}) = \{i = 1, \dots, p \mid f(\bar{x}) = f_i(\bar{x})\}$ tales que

$$0 = \sum_{i \in I(\bar{x})} \alpha_i \nabla f_i(\bar{x}), \quad \sum_{i \in I(\bar{x})} \alpha_i = 1.$$

A continuación veremos las condiciones necesarias y suficiente de Karush-Kuhn-Tucker que es de suma importancia el desarrollo del método FDIPA que se desarrollará en el capítulo siguiente.

2.4. Condiciones de Karush-Kuhn-Tucker

Los siguientes teoremas muestran la condición de primer y segundo orden de optimalidad, para un problema de optimización con restricciones de igualdad y desigualdad, cuyas demostraciones pueden ser encontradas en [4]. Estas condiciones de optimalidad llamadas condiciones de Karush-Kuhn-Tucker serán de gran utilidad para el desarrollo del método estudiado en el capítulo 3.

Teorema 2.16 (Primera condición de Karush-Kuhn-Tucker). *Sea D un subconjunto abierto no vacío de \mathbb{R}^n , y sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ para $i = 1, \dots, m$ y $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$ para $i = 1, \dots, l$. Consideremos el problema primal*

$$(K) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s. a} & g_i(x) \leq 0 \quad \text{para } i = 1, \dots, m \\ & h_j(x) = 0 \quad \text{para } j = 1, \dots, l \\ & x \in D \end{cases}$$

Sea \bar{x} una solución, y sea $I = \{i \mid g_i(\bar{x}) = 0\}$. Supongamos que f y g_i para $i \in I$ son diferenciables en \bar{x} , cada g_i para $i \notin I$ son continuas en \bar{x} , y que cada h_j para $j = 1, \dots, l$ son continuamente diferenciables en \bar{x} . Además que $\nabla g_i(\bar{x})$ para $i \in I$ y $\nabla h_j(\bar{x})$ para $j = 1, \dots, l$ son linealmente independientes. Si \bar{x} es solución local de (K), entonces existen escalares u_i para $i \in I$ y ν_j para $j = 1, \dots, l$ tal que,

$$\begin{aligned} \nabla f(\bar{x}) + \sum_{i=1}^m u_i \nabla g_i(\bar{x}) + \sum_{j=1}^l \nu_j \nabla h_j(\bar{x}) &= 0 \\ \mu_i g_i(\bar{x}) &= 0 && \text{para } i = 1, \dots, m \\ \mu_i &\geq 0 && \text{para } i = 1, \dots, m. \end{aligned} \quad (2.3)$$

Las condiciones que satisfacen la ecuación (2.3) son llamadas condiciones de optimalidad de Karush-Kuhn-Tucker y los puntos \bar{x} que lo satisfacen son llamados puntos KKT.

Definición 2.22 (Función lagrangiana). Considere el problema primal:

$$(P) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s.a} & x \in D \end{cases}$$

Donde $D = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m, h_j(x) = 0, j = 1, \dots, l\}$. Definimos la función Lagrangiana $L : \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^l \rightarrow \mathbb{R}$ definido por

$$L(x, \mu, \nu) = f(x) + \sum_{i=1}^m \mu_i g_i(x) + \sum_{j=1}^l \nu_j h_j(x).$$

Aquí, los μ_i y ν_j son llamados multiplicadores de Lagrange. Fijando los valores $\bar{\mu}$ y $\bar{\nu}$ definimos la restricción de la función lagrangiana, dada por

$$\phi(x) = L(x, \bar{\mu}, \bar{\nu}) = f(x) + \sum_{i=1}^m \bar{\mu}_i g_i(x) + \sum_{j=1}^l \bar{\nu}_j h_j(x).$$

Teorema 2.17. Consideremos el problema (K) definido en el teorema 2.16, donde la función objetivo y las restricciones son dos veces diferenciables, y donde D es un conjunto abierto. Sea \bar{x} un punto KKT del problema (K), con multiplicadores de Lagrange \bar{u} y \bar{v} asociados a las restricciones de desigualdad e igualdad, respectivamente. Sea $I = \{i \mid g_i(\bar{x}) = 0\}$, y denotemos $I^+ = \{i \in I \mid \bar{\mu}_i > 0\}$ y $I^0 = \{i \in I \mid \bar{\mu}_i = 0\}$. Con la definición de la restricción de la función lagrangiana $\phi(x)$ denotemos la hessiana en \bar{x} por

$$\nabla^2 \phi(\bar{x}) = \nabla^2 f(\bar{x}) + \sum_{i \in I} \bar{\mu}_i \nabla^2 g_i(\bar{x}) + \sum_{j=1}^l \bar{v}_j \nabla^2 h_j(\bar{x}),$$

donde $\nabla^2 f(\bar{x})$, $\nabla^2 g_i(\bar{x})$ para $i \in I$, y $\nabla^2 h_j(\bar{x})$ para $j = 1, \dots, l$, son las hessianas de f , g_i para $i \in I$, y h_j para $j = 1, \dots, l$, respectivamente, todos evaluados en \bar{x} . Se define el cono

$$C = \left\{ d \neq 0 \left| \begin{array}{ll} \nabla g_i(\bar{x})^t d = 0 & \text{para } i \in I^+, \nabla g_i(\bar{x})^t d \leq 0 \text{ para } i \in I^0, \\ \nabla h_i(\bar{x})^t d = 0 & \text{para todo } i = 1, \dots, l \end{array} \right. \right\}$$

Entonces si $d^t \nabla^2 \Phi(\bar{x}) d > 0$ para todo $d \in C$, tenemos en \bar{x} un mínimo local estricto para el problema (K).

Finalmente terminaremos este capítulo con el método Newton. Este método genera una secuencias de puntos $\{x^k\}$ que convergen a una solución del problema $\Phi(x) = 0$, donde $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$, de manera cuadrática.

2.5. Método de Newton para ecuaciones

El método de Newton clásico es diseñado para determinar un $\bar{x} \in \mathbb{R}^n$ tal que

$$\Phi(\bar{x}) = 0 \tag{2.4}$$

donde $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es al menos diferenciable. Sea $x^k \in \mathbb{R}^n$ una aproximación de alguna solución \bar{x} de la ecuación (2.4). Alrededor de x^k , podemos aproximar (2.4) por la linealización

$$\Phi(x^k) + J\Phi(x^k)(\bar{x} - x^k) = 0 \tag{2.5}$$

La ecuación (2.5) es llamada ecuación de iteración del método de Newton. La secuencia generada $\{x^k\}$ en el sistema lineal converge al punto \bar{x} .

Algoritmo 2.5.1 (Algoritmo del Método de Newton).

Escoger $x^0 \in \mathbb{R}^n$ y tomar $k = 0$.

1. *Calcule $x^{k+1} \in \mathbb{R}^n$ como una solución de la ecuación lineal (2.5), esto es*

$$x^{k+1} = x^k - (J\Phi(x^k))^{-1}\Phi(x^k)$$

2. *Tome $k = k + 1$ y volver al paso 1.*

Teorema 2.18 (Convergencia local del método de Newton). *Sea $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ una función diferenciable en una vecindad del punto $\bar{x} \in \mathbb{R}^n$, con derivada continua en ese punto. Si \bar{x} es una solución de la ecuación (2.4), tal que $J\Phi(\bar{x})$ sea no singular. Entonces para cualquier punto inicial $x^0 \in \mathbb{R}^n$ suficientemente próximo a \bar{x} , el algoritmo (2.5.1) genera una secuencia $\{x^k\}$ que está bien definida y que converge a \bar{x} . Ella converge de forma superlineal, y si la derivada de Φ es Lipschitz en una vecindad de \bar{x} , entonces la tasa de convergencia es cuadrática.*

Para más detalles de la convergencia del método de Newton e ilustraciones, pueden ver por ejemplo en [2], [3] y [4].

Capítulo 3

Método de Direcciones Factibles de Puntos Interiores para Optimización No Lineal Diferenciable

El método que presentaremos a continuación, fue propuesto por el Doctor en matemática José Herskovits [16], llamado FDIPA (Feasible Direction Interior Point Algorithm) es un método de puntos interiores. Es un método fácil de programar (y con convergencia superlineal), pues no es necesario resolver subproblemas cuadráticos, y tampoco se trata de métodos de penalidad, o de barrera, o filtros. El método genera una secuencia de direcciones de descenso para la función objetivo el cual solo requiere la solución de dos sistemas de ecuaciones lineales con una misma matriz principal, el cual siempre es invertible en cada iteración, seguida de una búsqueda lineal inexacta. La convergencia del método es global para aquellos puntos que son Karush-Kuhn-Tucker.

A pesar de ser un método para problemas de optimización diferenciable, este método es bien utilizado en problemas de programación no diferenciable. El cual será empleado en el método que desarrollaremos en el capítulo 5, el cual es un problema de optimización convexa no diferenciable. Una característica importante del método FDIPA es la facilidad de hallar una dirección de descenso y factible, que es determinado al resolver dos sistemas lineales generados por el método FDIPA y que en cada iteración siempre esta dirección es única.

3.1. Método FDIPA

El método que desarrollaremos (FDIPA) en este capítulo es muy importante para el estudio de muchos nuevos métodos ya realizados en estos tiempos. Los cuales son una modificación de los métodos Plano Cortantes, Bundles, u otros. Si bien estos métodos son sobre problemas de optimización no diferenciables, el FDIPA es usado como ayuda para la solución de estos nuevos métodos no diferenciables. El método Feasible Directions Interior Point Technique for Nonlinear Optimization(FDIPA) resuelve problemas de programación no lineal diferenciables con restricciones de desigualdad e igualdad diferenciables. En este capítulo solo desarrollaremos el caso donde las restricciones son de desigualdad.

Consideremos el problema de programación no lineal diferenciable con restricciones de desigualdad

$$(P) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s.a} & g(x) \leq 0 \end{cases}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $g = (g_1, \dots, g_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ son funciones dos veces diferenciables. Denotemos $Jg(x) \in \mathbb{R}^{n \times m}$ a la matriz Jacobiana de g , λ un vector en \mathbb{R}^m el multiplicador de Lagrange, $\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ el conjunto factible y Ω^0 su interior. Definiendo la función lagrangiana asociada al problema (P)

$$L(x, \lambda) = f(x) + \lambda^t g(x)$$

Cuya matriz hessiana es dada por

$$\Gamma(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x)$$

Luego, las condiciones de Karush-Kuhn-Tucker (KKT) de primer orden de optimalidad son dado por

$$\nabla f(x) + Jg(x)\lambda = 0 \tag{3.1}$$

$$G(x)\lambda = 0 \quad (3.2)$$

$$g(x) \leq 0 \quad (3.3)$$

y

$$\lambda \geq 0. \quad (3.4)$$

Donde $G(x) = \text{diag}(g_1(x), \dots, g_n(x))$ es una matriz diagonal con g_i , para $i = 1, \dots, m$, de elementos en la diagonal.

Se define un vector (x^*, λ^*) que satisface las igualdades (3.1) y (3.2) como un par estacionario del problema (P) y un vector KKT si además satisface las condiciones de KKT. Un vector x^* es un punto estacionario si existe un λ^* tal que (x^*, λ^*) es un par estacionario.

3.2. Hipótesis adicionales para la convergencia global

Para la convergencia global del método se supone, las siguientes hipótesis:

Hipótesis 2.1. Existe un número real a tal que el conjunto $\Omega_a = \{x \in \Omega \mid f(x) \leq a\}$ es compacto con interior Ω_a^0 no vacío.

Hipótesis 2.2 $x \in \Omega_a^0$ satisface $g(x) < 0$.

Hipótesis 2.3 Las funciones f y g son continuamente diferenciables en Ω_a y sus derivadas satisfacen la condición de Lipschitz.

Hipótesis 2.4 (Condición de regularidad) Para cada $x \in \Omega_a$ los vectores $\nabla g_i(x)$ son linealmente independientes, para cada i tal que $g_i(x) = 0$.

Iniciando con un punto interior, el algoritmo de FDIPA genera una secuencia $\{x^k\}_{k \in \mathbb{N}}$ de puntos interiores de tal manera que

$$f(x^{k+1}) < f(x^k) \quad y \quad g_i(x^k) < 0, \quad \forall i = 1, 2, \dots, m$$

y converge a un conjunto de puntos KKT del problema primal. En cada iteración es encontrada una dirección de búsqueda lineal y factible, es decir, una dirección de descenso para la función objetivo y una dirección factible para Ω^0 . Una búsqueda lineal es entonces realizada para garantizar que el nuevo punto se encuentre en el interior de

Ω y que también garantice la convergencia global.

De las ecuaciones (3.1) y (3.2) consideremos lo siguiente

$$Y = (x, \lambda) \quad y \quad \phi(Y) = \begin{bmatrix} \nabla f(x) + Jg(x)\lambda \\ G(x)\lambda \end{bmatrix}$$

con

$$J\phi(Y) = \begin{bmatrix} \Gamma(x, \lambda) & Jg(x) \\ \Lambda(Jg(x))^t & G(x) \end{bmatrix}$$

donde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$ es una matriz $m \times m$, y $\Gamma(x, \lambda)$, $G(x)$ y $Jg(x)$, son la matriz Hessiana de la función lagrangiana de f , la matriz diagonal con elementos en la diagonal $g_i(x)$, para $i = 1, \dots, m$ y la matriz Jacobiana de $g(x)$, respectivamente. Luego, una iteración de Newton es realizada para resolver el sistema de ecuaciones lineales $\phi(x, \lambda) = \phi(Y) = 0$, suponiendo que estamos en la iteración $Y^k = (x^k, \lambda^k)$, iteración k , se define el punto $Y = (x^{k+1}, \lambda^{k+1})$ que es la solución del sistema lineal

$$J\phi(Y^k)(Y - Y^k)^t = -(\phi(Y^k))^t$$

el cual puede ser escrito como

$$\begin{bmatrix} B^k & Jg(x^k) \\ \Lambda^k(Jg(x^k))^t & G(x^k) \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) + Jg(x^k)\lambda^k \\ G(x^k)\lambda^k \end{bmatrix} \quad (3.5)$$

donde $\Gamma(x^k, \lambda^k)$ es sustituido por B_k , y B_k puede ser tomado igual a una estimación quasi-Newton, ver por ejemplo [2] y [3], o por la matriz identidad. De hecho, como un requerimiento para la convergencia global del presente algoritmo, B_k debe ser una matriz simétrica definida positiva en cada iteración k .

En lo que sigue, se introduce alguna modificación en la iteración (3.5) de manera que al obtener un punto interior (x^k, λ^k) una nueva estimación haga decrecer a la función objetivo. Con este propósito, se define d_α^k en el espacio primal, como $d_\alpha^k = x^{k+1} - x^k$ y $\lambda_\alpha^k = \lambda^{k+1}$.

Entonces, de (3.5) obtenemos:

$$B_k d_\alpha + Jg(x^k)\lambda_\alpha^k = -\nabla f(x^k) \quad (3.6)$$

y

$$\Lambda^k (Jg(x^k))^t d_\alpha^k + G(x^k)\lambda_\alpha^k = 0, \quad (3.7)$$

Donde la solución $(d_\alpha^k, \lambda_\alpha^k)$ de este sistema proporciona una dirección de descenso d_α^k y una nueva estimación de λ_α^k para λ . En [16], Herskovits prueba que d_α^k es una dirección de descenso para f y es cero en un punto estacionario y por lo tanto una solución del problema es dado. No obstante, d_α^k puede no ser una dirección factible, pues cuando alguna restricción se aproxima a cero, d_α^k tiende a una dirección tangente del conjunto factible.

En efecto, reescribiendo la ecuación (3.7) se tiene

$$\lambda_i^k (\nabla g_i(x^k))^t \cdot d_\alpha^k + g_i(x^k)\lambda_{\alpha_i}^k = 0, \quad i = 0, 1, \dots, m \quad (3.8)$$

lo que implica que $(\nabla g_i(x^k))^t \cdot d_\alpha^k = 0$ para todo i , tal que $g_i(x^k) = 0$. Para modificar esta dirección, se define un nuevo sistema lineal en d^k y $\bar{\lambda}^k$, tal que se incrementa en el lado derecho de (3.7) el vector $-\rho^k \Lambda^k$ (donde ρ^k es llamado parámetro de deflaxión), esto es,

$$B_k d^k + Jg(x^k)\bar{\lambda}^k = -\nabla f(x^k),$$

$$\Lambda^k (Jg(x^k))^t d^k + G(x^k)\bar{\lambda}^k = -\rho^k \Lambda^k \quad (3.9)$$

donde $\rho^k > 0$, d^k es una nueva dirección y $\bar{\lambda}^k$ es una nueva estimación de λ . En este caso, (3.9) es equivalente

$$\lambda_i^k (\nabla g_i(x^k))^t \cdot d^k + g_i(x^k)\bar{\lambda}_i^k = -\rho^k \lambda_i^k \quad i = 1, 2, \dots, m$$

y

$$(\nabla g_i(x^k))^t \cdot d^k = -\rho^k < 0$$

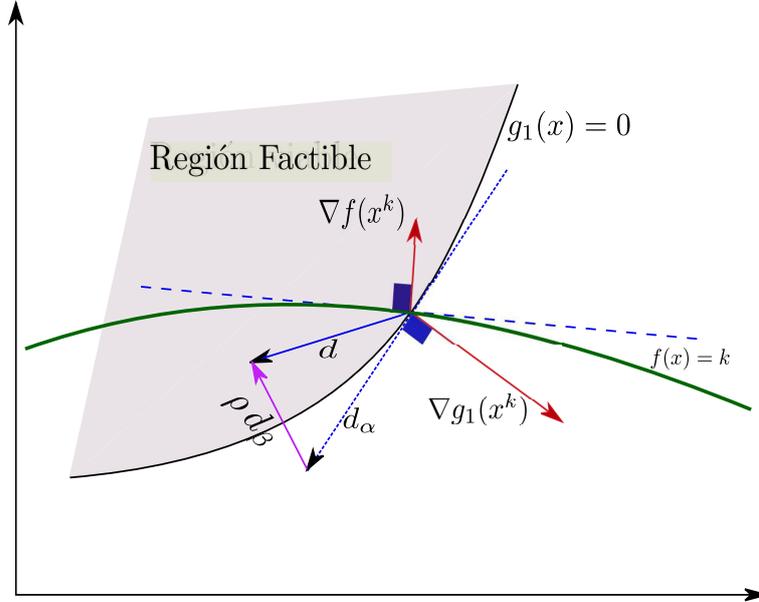


Figura 3.1: d dirección factible

para una restricción activa tal que $\lambda_i^k \neq 0$, y en este caso tendremos que d^k es una dirección factible.

La decisión de tomar un número negativo del lado derecho de la ecuación (3.7) es que aquella produce una deflexión en la dirección en d^k el cual es proporcional a ρ^k , en dirección a la región factible. Para garantizar que efectivamente d^k sea una dirección de búsqueda, ρ^k debe ser escogido de manera conveniente, para este resultado puede ver [16].

Finalmente notemos que d^k puede ser obtenido resolviendo dos sistemas lineales,

$$\begin{aligned} B_k d_\alpha^k + Jg(x^k) \lambda_\alpha^k &= -\nabla f(x^k) \\ \Lambda^k (Jg(x^k))^t d_\alpha^k + G(x^k) \lambda_\alpha^k &= 0 \end{aligned}$$

y

$$\begin{aligned} B_k d_\beta^k + Jg(x^k) \lambda_\beta^k &= 0 \\ \Lambda^k (Jg(x^k))^t d_\beta^k + G(x^k) \lambda_\beta^k &= -\rho^k \Lambda^k, \end{aligned}$$

de modo que

$$d^k = d_\alpha^k + \rho^k d_\beta^k$$

$$\bar{\lambda}^k = \lambda_\alpha^k + \rho^k \lambda_\beta^k.$$

3.3. Algoritmo de Puntos Interiores de Direcciones Factibles

En esta sección presentamos el algoritmo FDIPA, resultado del método FDIPA desarrollado en la sección anterior.

Algoritmo 3.3.1.

Parámetros $\xi \in \langle 0, 1 \rangle$, $\eta \in \langle 0, 1 \rangle$, $\varphi > 0$ y $\nu \in \langle 0, 1 \rangle$.

Datos iniciales. $x^0 \in \Omega_\alpha^0$, $0 < \lambda^0 \in \mathbb{R}^m$, $B_k \in \mathbb{R}^{n \times n}$ matriz simétrica definida positiva y $k=0$.

Paso 1. Cálculo de una dirección de búsqueda.

(i) Calcule $(d_\alpha^k, \lambda_\alpha^k)$ resolviendo el siguiente sistema lineal

$$B_k d_\alpha^k + Jg(x^k) \lambda_\alpha^k = -\nabla f(x^k), \quad (3.10)$$

$$\Lambda^k (Jg(x^k))^t d_\alpha^k + G(x^k) \lambda_\alpha^k = 0. \quad (3.11)$$

Si $d_\alpha^k = 0$, parar.

(ii) Calcule $(d_\beta^k, \lambda_\alpha^k)$ resolviendo el sistema lineal

$$B_k d_\beta^k + Jg(x^k) \lambda_\beta^k = 0, \quad (3.12)$$

$$\Lambda^k (Jg(x^k))^t d_\beta^k + G(x^k) \lambda_\beta^k = -\Lambda^k. \quad (3.13)$$

(iii) Si $(d_\beta^k)^t \nabla f(x^k) > 0$, tomamos

$$\rho^k = \min \left[\varphi \|d_\alpha^k\|_2^2; (\xi - 1) \frac{(d_\alpha^k)^t \cdot \nabla f(x^k)}{(d_\beta^k)^t \cdot \nabla f(x^k)} \right]. \quad (3.14)$$

Otro caso, tomamos

$$\rho^k = \varphi \|d_\alpha^k\|_2^2. \quad (3.15)$$

(iv) Calcule la dirección de búsqueda lineal

$$d^k = d_\alpha^k + \rho^k d_\beta^k \quad (3.16)$$

y también

$$\bar{\lambda}^k = \lambda_\alpha^k + \rho^k \lambda_\beta^k \quad (3.17)$$

Paso 2. Búsqueda lineal

Calcule t , el primer número de la secuencia $\{1, \nu, \nu^2, \nu^3, \dots\}$ satisfaciendo

$$f(x^k + td^k) \leq f(x^k) + t\eta(\nabla f(x^k))^t \cdot d^k \quad (3.18)$$

y

$$g_i(x^k + td^k) < 0, \text{ si } \bar{\lambda}_i^k \geq 0, \quad (3.19)$$

o

$$g_i(x^k + td^k) \leq g_i(x^k) \quad (3.20)$$

otro caso.

Paso 3. Actualizando

(i) Tomando

$$x^{k+1} = x^k + td^k$$

actualizar $k = k + 1$ y

$$\lambda^k > 0 \text{ y } B_k$$

Note que el algoritmo usa la regla de Armijo para la búsqueda lineal con una cierta variante en la restricción del conjunto factible. En la actualización de B y λ , el algoritmo usa diferentes reglas de actualización, para estas reglas de actualización ver [16].

3.4. Convergencia del algoritmo FDIPA

Para la convergencia del algoritmo es necesario algunas hipótesis adicionales al método.

Hipótesis 3.1 Existen números positivos λ^I, λ^S y \bar{g} tales que $0 < \lambda_i \leq \lambda^S, i = 1, \dots, m$ y $\lambda_i \geq \lambda^I$ para cualquier i tal que $g_i(x) \geq -\bar{g}$.

Hipótesis 3.2 Existen números reales positivos δ_1 y δ_2 tales que

$$\delta_1 \|d\|_2^2 \leq d^T B d \leq \delta_2 \|d\|_2^2, \text{ para cualquier } d \in \mathbb{R}^n.$$

La convergencia del algoritmo y de otras versiones de este pueden ser encontradas en [16]. El siguiente lema muestra que los sistemas lineales dados por el algoritmo tiene solución y además, es única y por las hipótesis adicionales dan lugar a que $\lambda_\alpha, \lambda_\beta, d_\alpha$ y d_β sean valores finitos.

Lema 3.1. *Dado cualquier $x \in \Omega_a$, para cualquier matriz simétrica definida positiva B y cualquier $\lambda \in \mathbb{R}^m$ tal que $\lambda_i > 0$ con $g_i(x) = 0$, se tiene que la matriz*

$$M(x, \lambda, B) = \begin{bmatrix} B & Jg(x) \\ \Lambda(Jg(x))^t & G(x) \end{bmatrix}$$

es no singular.

Prueba: Para la verificación de la no singularidad de $M(x, \lambda, B)$ es suficiente demostrar que el sistemas lineal homogéneo

$$M(x, \lambda, B)(d, \mu) = 0$$

tiene solución única. En efecto, del sistema anterior tenemos

$$Bd + Jg(x)\mu = 0 \tag{3.21}$$

$$\Lambda(Jg(x))^t d + G(x)\mu = 0 \tag{3.22}$$

Denotemos $I_m = \{1, \dots, m\}$, luego,

$$I_m = I(x) \cup \{i \in I_m - I(x) \mid \lambda_i = 0\} \cup \{i \in I_m - I(x) \mid \lambda_i > 0\}$$

donde $I(x) = \{i \in I_m \mid g_i(x) = 0, \lambda_i > 0\}$ y se da la regularidad de los $g_i(x)$. Si $i \in I(x)$ entonces por hipótesis del problema $\lambda_i > 0$ y de (3.22),

$$(\nabla g_i(x))^t d = 0.$$

Si $i \in \{i \in I_m - I(x) \mid \lambda_i = 0\}$ entonces $g_i(x) \neq 0$, y de (3.22),

$$\mu_i = 0.$$

Si $i \in \{i \in I_m - I(x) \mid \lambda_i > 0\}$ entonces

$$(\nabla g_i(x))^t d = -\frac{\mu_i}{\lambda_i} g_i(x),$$

Multiplicando (3.21) por d^t

$$\begin{aligned} d^t B d + d^t J g(x) \mu &= 0 \\ d^t B d + \sum_{i \in \{i \in I_m - I(x) \mid \lambda_i > 0\}} -\frac{\mu_i^2 d^t}{\lambda_i} \nabla g_i(x) &= 0 \end{aligned}$$

y $g_i(x) < 0$ ya que $x \in \Omega_a$. Luego de la última ecuación $d = 0$ y $\mu_i = 0$ para $i \in \{i \in I_m - I(x) \mid \lambda_i > 0\}$.

Reemplazando estos valores en (3.21)

$$\sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0,$$

y por la condición de regularidad $\mu_i = 0$ para todo $i \in I(x)$. Así $(d, \mu) = (0, 0)$. \square

El siguiente lema, cuya demostración se encuentra en [16], muestra que la dirección d , dado por el algoritmo FDIPA, es una dirección de descenso.

Lema 3.2. *La dirección de búsqueda d satisface*

$$d^t \cdot \nabla f(x) \leq \xi (d_\alpha)^t \cdot f(x).$$

Las demostraciones de los siguientes resultados que muestran la convergencia del algoritmo FDIPA pueden ser encontrados en [15] y [16].

Proposición 3.1. *Sea $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continuamente diferenciable y tal que $\nabla\phi$ satisface la condición de Lipschitz en un subconjunto $\Gamma \subseteq \mathbb{R}^n$, es decir, existe un constante positiva k tal que para cualquier $x, y \in \Gamma$,*

$$\|\nabla\phi(y) - \nabla\phi(x)\|_2 \leq k\|y - x\|_2,$$

entonces, si $[x, y] \subseteq \Gamma$ la siguiente condición es satisfecha:

$$\phi(y) \leq \phi(x) + (y - x)^t \cdot \nabla\phi(x) + k\|y - x\|_2^2$$

El lema siguiente garantiza la existencia de la longitud de paso por el algoritmo FDIPA.

Lema 3.3. *Existe $\tau > 0$ tal que para cualquier $x \in \Omega_a$ y d calculado en el paso 1 del algoritmo, las condiciones (3.18)-(3.20) son verificadas, para cualquier $t \in [0, \tau]$.*

Lema 3.4. *Cualquier punto de acumulación x^* de la secuencia $\{x^k\}$, generada por el algoritmo FDIPA, es un punto estacionario. Además, $(x^*, \lambda_\alpha(x^*))$ constituyen un par estacionario.*

El siguiente teorema muestra que un punto de acumulación de la secuencia $\{x^k\}$ generada por el algoritmo FDIPA converge a la solución del problema (P).

Teorema 3.1. *Cualquier punto de acumulación x^* de la secuencia $\{x^k\}$ generada por el algoritmo es un punto Karush-Kuhn-Tucker del problema (P).*

Capítulo 4

Métodos para programación convexas no diferenciables

Dedicamos este capítulo al estudio de programación convexa no diferenciable. Como veremos a continuación para los métodos de programación con función objetivo no diferenciable es necesario desarrollar otras técnicas computacionales, ya que, a diferencia de los métodos de programación diferenciable en ventas a los métodos no diferenciables, poseen un criterio de parada que son muy eficaces computacionalmente, como por ejemplo el criterio de la gradiente el cual consiste en tener $\|\nabla f(x)\| > \epsilon$, para algún $\epsilon > 0$ fijo. Este criterio no tiene sentido para métodos de programación no diferenciables. Otra desventaja para los métodos de programación diferenciables es que podemos encontrar una dirección de descenso (como por ejemplo $-\nabla f(x)$), mientras en los métodos de programación no diferenciables escogiendo como dirección $s \in \partial f(x)$ ella no siempre será una dirección de descenso, aunque pueda ser una dirección factible, y por lo tanto la búsqueda lineal no siempre tendrá sentido. En este capítulo nos dedicaremos a ver cuales son los caminos apropiados para subsanar dichas desventajas. No obstante los métodos que presentamos en este capítulo son los más conocidos en la literatura y los cuales son: el método del Subgradiente, método de Planos Cortantes y método Bundle. Este último, método de Bundle, a diferencia del método de Planos Cortantes posee un criterio de parada y un centro estabilizador el cual garantiza que la siguiente iteración esté más cerca a la solución óptima. Además, un control en la inserción de planos de cortes. Es por ello que pueden ser de gran ayuda para mejorar

la parte computacional del método que desarrollaremos en el siguiente capítulo.

4.1. Método del Subgradiente

Consideremos el siguiente problema de optimización sin restricciones

$$(P) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s.a} & x \in \mathbb{R}^n \end{cases}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa sobre \mathbb{R}^n no necesariamente diferenciable.

El método que empezamos a describir es llamado método del Gradiente Generalizados, detallados por muchos autores, ver por ejemplo [2], [4] y [25], reemplaza la condición de dirección negativa del gradiente por una dirección negativa basada en un subgradiente, es decir,

$$d^k = -s^k, \text{ donde } s^k \in \partial f(x^k).$$

Esta dirección, a diferencia de la dirección del gradiente de una función diferenciable no necesariamente es una dirección de descenso, aunque da lugar a una nueva iteración acercándose cada vez más a la solución óptima con una longitud de paso adecuada, que en muchos casos es suficientemente pequeño.

Es por ello que el método del Subgradiente es considerado como método estructuralmente simple. Esta presenta muchas veces muy buenos resultados numérico debido a la manera de cómo se escoge un subgradiente de entre muchos. Existen muchas variaciones de este método descritos por muchos autores y de sus aplicaciones, ver por ejemplo [13] y [5].

Presentamos ahora una versión del algoritmo del método del Subgradiente.

Algoritmo 4.1.1 (Algoritmo del método del Subgradiente).

Paso 1. Tome cualquier punto $x^0 \in \mathbb{R}^n$, y tome $k = 0$

Paso 2. Genere una iteración $x^{k+1} = x^k - t^k s^k$, con $s^k \in \partial f(x^k)$.

Paso 3. Tome $k=k+1$, e ir a el paso 2.

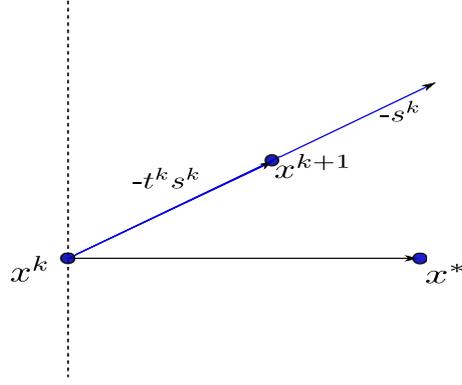


Figura 4.1: **Condición para escoger la longitud de paso para el método del Subgradiente**

Condición para escoger la longitud de paso

En esta sección estableceremos una manera de escoger la longitud de paso. El cual nos garantiza la convergencia global del método del Subgradiente. En este caso si escogemos una longitud de paso muy grande puede generar en el algoritmo una secuencia $\{x^k\}$ que oscile innecesariamente alrededor de la solución óptima. En cambio, si escogemos longitudes de paso muy pequeños, el algoritmo se aproximará muy lentamente a la solución óptima, vea por ejemplo [13] y [5]. Es por ello que presentamos una manera de cómo podemos escoger la longitud de paso de manera apropiada que evita estos problemas.

Supongamos que x^* sea solución del problema (P). Tomando un $s^k \in \partial f(x^k)$, entonces

$$\begin{aligned} f(x) - f(x^k) &\geq (s^k)^t \cdot (x - x^k), \quad \forall x \in \mathbb{R}^n \\ f(x^*) - f(x^k) &\leq (s^k)^t \cdot (x^* - x^k) \end{aligned},$$

del cual se tiene

$$-(s^k)^t \cdot (x^* - x^k) \geq f(x^k) - f(x^*),$$

dado que x^* es solución del problema (P), se tiene que $f(x^k) - f(x^*) \geq 0$ para todo $k \in \mathbb{N}$. Por lo tanto, $-(s^k)^t \cdot (x^* - x^k) \geq 0$ y consecuentemente el ángulo formado entre $-s^k$ y $x^* - x^k$ será agudo, tal como muestra la figura 4.1.

Para una longitud de paso $0 < t^k < 2 \frac{f(x^k) - f(x^*)}{\|s^k\|^2}$, es fácil de ver que:

$$\|x^{k+1} - x^*\| < \|x^k - x^*\|.$$

Así, se tiene que x^{k+1} está cada vez más próximo a la solución x^* que la iteración x^k . El siguiente teorema demuestra nuestra afirmación dada y cuya prueba puede ser encontrada en [24].

Teorema 4.1. *Sea x^* una solución del problema (P), entonces la secuencia $\{x^k\}$ generada por el algoritmo del subgradiente es de orden lineal, es decir,*

$$\|x^{k+1} - x^*\| < \|x^k - x^*\|$$

siempre que

$$0 < t^k < 2 \frac{f(x^k) - f(x^*)}{\|s^k\|^2}.$$

Una contrariedad de este teorema es que debemos conocer el valor óptimo del problema y por lo tanto no podríamos acotar t^k como se muestra en el teorema. Por lo tanto no se garantiza la convergencia lineal. Es por ello que debemos condicionar t^k de otra manera, tal que se garantice la convergencia, sin tener presente el valor óptimo x^* . Para esta finalidad veamos:

Tomando un punto inicial x^0

$$\|x^k - x^0\| \leq \|x^k - x^{k-1}\| + \|x^{k-1} - x^{k-2}\| + \dots + \|x^2 - x^1\| + \|x^1 - x^0\|$$

como $x^{k+1} = x^k - t^k s^k$, entonces

$$\begin{aligned} \|x^k - x^0\| &\leq t^{k-1} + t^{k-2} + \dots + t^0 = \sum_{j=0}^{k-1} t^j \\ &\leq r \end{aligned}$$

Por lo tanto, el método encontrará una solución x^* si y solamente si $\forall k \in \mathbb{N}$ se tiene $x^* \in B(x^0, r)$, donde x^0 representa el punto inicial. Para contornar este problema, se impone la condición para $\{t^k\}$ de manera que $\sum_{k=0}^{\infty} t^k = +\infty$, es decir, que sea lo suficientemente grande para tomar dicho punto inicial.

El próximo teorema establece la convergencia del método del Subgradiente sin conocer la solución óptima.

Teorema 4.2 (Convergencia del método del Subgradiente). *Supongamos que el al-*

goritmo 4.1.1, las secuencias $\{t^k\}$ y $\{s^k\}$ satisfacen las condiciones $\lim_{k \rightarrow \infty} t^k = 0$ y $\sum_{k=0}^{\infty} t^k = +\infty$. Entonces, para la secuencia $\{x^k\}$ generada por el algoritmo 4.1.1, se tiene

$$\liminf_{k \rightarrow \infty} f(x^k) = \bar{v}$$

donde $\bar{v} = \inf_{x \in \mathbb{R}^n} f(x)$. Si además suponemos que el conjunto X^* de las soluciones del problema (P) es no vacío y acotado, entonces $\{x^k\}$ es una secuencia acotada y cada punto de acumulación de la secuencia pertenece al conjunto X^* .

Prueba:

$$\begin{aligned} \|x^{k+1} - x\|^2 &= \|x^k - x\|^2 + 2\langle x^{k+1} - x^k, x^k - x \rangle + \|x^{k+1} - x^k\|^2 \\ &= \|x^k - x\|^2 + 2\langle -t^k s^k, x^k - x \rangle + (t^k)^2 \\ &= \|x^k - x\|^2 + 2\langle t^k s^k, x - x^k \rangle + (t^k)^2 \\ &\leq \|x^k - x\|^2 + 2t^k (f(x) - f(x^k)) + (t^k)^2 \end{aligned} \tag{4.1}$$

Supongamos que $\liminf_{k \rightarrow \infty} f(x^k) > \bar{v}$. En este caso existe un $x \in \mathbb{R}^n$, un número $\delta > 0$ y un índice k_1 , tales que

$$f(x) < f(x^k) - \delta, \quad \forall k \geq k_1.$$

Además $\lim_{k \rightarrow \infty} t^k = 0$, existe k_2 tal que $t^k \leq \delta$, $\forall k \geq k_2$, tomando $\bar{k} = \max\{k_1, k_2\}$ y de la ecuación (4.1)

$$\begin{aligned} \|x^{k+1} - x\|^2 &\leq \|x^k - x\|^2 + t^k (t^k - 2\delta) \\ &\leq \|x^k - x\|^2 + t^k (\delta - 2\delta) = \|x^k - x\|^2 - \delta t^k, \quad \forall k \geq \bar{k} \end{aligned} \tag{4.2}$$

Por lo tanto

$$\begin{aligned} \delta \sum_{i=\bar{k}}^k t^i &\leq \sum_{i=\bar{k}}^k (\|x^i - x\|^2 - \|x^{i+1} - x\|^2) \\ &= \|x^{\bar{k}} - x\|^2 - \|x^{k+1} - x\|^2 \\ &\leq \|x^{\bar{k}} - x\|^2 \end{aligned}$$

el cual es una contradicción. □

Una condición más fuerte sobre la convergencia del método del Subgradiente es dado por el siguiente teorema.

Teorema 4.3. *Supongamos que el problema (P) tenga una solución. Supongamos que en el algoritmo 4.1.1, $t^k = \frac{\beta_k}{\|s^k\|}$, $\forall k$, y que la secuencia $\{\beta_k\}$ satisface las condiciones*

$$\sum_{k=0}^{\infty} \beta_k = +\infty, \quad \sum_{k=0}^{\infty} \beta_k^2 < +\infty \quad (4.3)$$

Entonces toda secuencia $\{x^k\}$ generada por el algoritmo 4.1.1 converge a una solución del problema (P).

Prueba: *Sea $\bar{x} \in \mathbb{R}^n$ una solución del problema (P). Sabemos que, $f(\bar{x}) \leq f(x^k), \forall k$. Tomando en la ecuación (4.1) $x = \bar{x}$, y teniendo en cuenta $t^k = \frac{\beta_k}{\|s^k\|}$, obtenemos que*

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 + \beta_k^2.$$

Sea k arbitrario pero fijo. Utilizando la última desigualdad en cadena, para cualquier $j \geq k + 1$ con k pero arbitrario, obtenemos

$$\begin{aligned} \|x^j - \bar{x}\|^2 &\leq \|x^k - \bar{x}\|^2 + \sum_{i=k}^{j-1} \beta_i^2 \\ &\leq \|x^k - \bar{x}\|^2 + \sum_{i=0}^{\infty} \beta_i^2 < +\infty, \end{aligned} \quad (4.4)$$

por la segunda relación en (4.3), concluimos que la secuencia $\{x^k\}$ es acotada. Por tanto, la secuencia $\{s^k\}$ también es acotada, ya que $s^k \in \partial f(x^k)$ para todo k . De $t^k = \frac{\beta_k}{\|s^k\|}$ y de la condición (4.3), obtenemos que $\sum_{k=0}^{\infty} t^k = +\infty$ y $\lim_{k \rightarrow \infty} t^k \|d^k\|^2 = \beta^k = 0$. Luego, por el teorema anterior, $\liminf_{k \rightarrow \infty} f(x^k) = f(\bar{x})$. Por la continuidad de f y desde que $\{x^k\}$ es acotada y por tanto posee un punto acumulación $\hat{x} \in \mathbb{R}^n$ tal que $f(\hat{x}) = f(\bar{x})$, es decir, \hat{x} es una solución del problema (P). Podemos entonces tomar $\bar{x} = \hat{x}$ en la ecuación (4.4). Tomando $j \geq k + 1$,

$$\|x^j - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + \sum_{i=k}^{\infty} \beta_i^2, \quad (4.5)$$

donde de (4.3),

$$\lim_{k \rightarrow \infty} \left(\sum_{i=k}^{\infty} \beta_i^2 \right) = 0.$$

En particular, para todo $\delta > 0$ arbitrariamente pequeño podemos escoger k suficientemente grande, tal que

$$\frac{\delta}{2} > \sum_{i=k}^{\infty} \beta_i^2.$$

Como \hat{x} es un punto de acumulación de la secuencia $\{x^k\}$, podemos también escoger un k tal que

$$\frac{\delta}{2} > \|x^k - \hat{x}\|^2.$$

De (4.5), concluimos que para todo $\delta > 0$, existe k tal que

$$\delta > \|x^j - \hat{x}\|^2, \quad \forall j \geq k + 1.$$

Finalmente, obtenemos que la secuencia $\{x^k\}$ converge a \hat{x} , y es una solución del problema (P).

□

4.1.1. Deficiencias del método del Subgradiente

Debido a que el método del Subgradiente no posee una regla de parada eficiente, esto es, si consideramos que $\|s^k\| > \epsilon$, para $\epsilon > 0$, no se puede esperar que ella genere un descenso de la función objetivo, aunque podríamos usar el teorema 2.15 que garantiza que si x^* es un mínimo del problema (P) es necesario y suficiente que cumpla que $0 \in \partial f(x^*)$, pero implicaría encontrar una secuencia $\{s^k\}$ de subgradientes en $\partial f(x^k)$ de manera que ella converja a 0 en al menos en un punto de acumulación y que a su vez x^k converja a x^* y con ello obtener un criterio de parada.

Sabemos que por el teorema 2.10 nos garantiza una condición necesaria y suficiente para encontrar una dirección d de descenso, esto es,

$$d \in D_f(x) \iff \langle y, d \rangle < 0, \quad \forall y \in \partial f(x).$$

Pero esto implicaría hacer la comparación entre todas las subgradientes de f en el punto x , el cual es imposible computacionalmente.

Sabemos para una función diferenciable f que $d = -\nabla f(x)$ es una dirección de des-

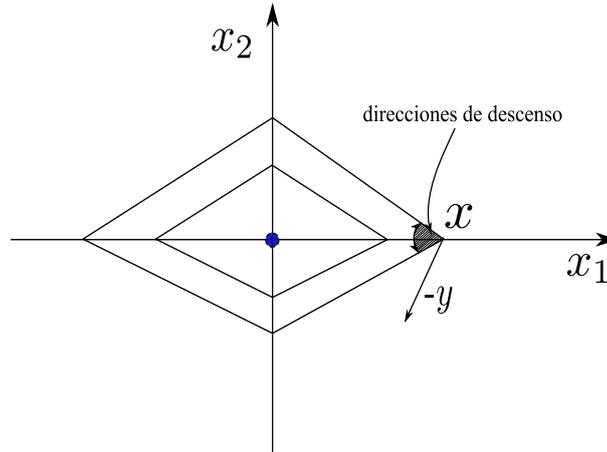


Figura 4.2: **La dirección $y = -(1, 2)$ que es un anti-subgradiente de f en x pero no es una dirección de descenso de f .**

censo, pero no es tan cierto para una función f no diferenciable, es decir, si escogemos $-s$ de $s \in \partial f(x)$ esto no es necesariamente una dirección de descenso como muestra el siguiente ejemplo:

Sea $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ una función definida por $f(x_1, x_2) = |x_1| + 2|x_2|$, y consideremos un punto $x = (x_1, 0)$, donde $x_1 > 0$, es arbitrario. Es fácil verificar que

$$(1, 2) = \bar{y} \in \partial f(x_1, 0),$$

y para todo $t > 0$ suficientemente pequeño, $f(\bar{x} - t\bar{y}) = x_1 + 3t > f(x_1, 0)$ lo que implica que

$$-\bar{y} \notin D_f(x_1, 0)$$

En la figura 4.2 se muestra que el anti-subgradiente $(-s)$ puede no ser una dirección de búsqueda.

Los siguientes métodos que trataremos, los cuales son más estable con respecto al método del Subgradiente, son llamados los método Planos Cortantes y método Bundle, que resultan de una aproximación cuasilineal a la función objetivo. Vamos a ver que a diferencia del método Planos Cortantes el método Bundle tiene un control de estabilización en cada iteración dada, es decir, en cada iteración garantiza el decrecimiento de la función objetivo.

4.2. Método de Planos Cortantes

Considere el siguiente problema de optimización

$$(C) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s.a} & x \in D \end{cases}$$

donde D es un conjunto convexo y compacto, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa en \mathbb{R}^n el cual no necesariamente es diferenciable.

La idea principal del método de Planos Cortantes es utilizar las informaciones ya acumuladas a lo largo de las iteraciones anteriores para determinar una nueva aproximación (inferior) de la función f , el cual es lineal, cada vez más cercana a la función. Considerando que en el inicio de la iteración con índice $n + 1$, se tiene la siguiente información, $x^k \in \mathbb{R}^n$, $s^k \in \partial f(x^k)$, $k = 0, 1, 2, \dots, n$. La siguiente iteración de los método de Planos Cortantes consiste en la resolución del problema

$$(C_n) \quad \begin{cases} \text{minimizar} & \psi_n(x) \\ \text{s.a} & x \in D, \end{cases}$$

donde $\psi_n : \mathbb{R}^n \rightarrow \mathbb{R}$, $\psi_n(x) = \max_{k=0,1,\dots,n} \{f(x^k) + (s^k)^t \cdot (x - x^k)\}$. Para su solución este problema puede ser escrito en su forma equivalente

$$\begin{cases} \text{minimizar} & t \\ \text{s.a} & (t, x) \in U_n \end{cases}$$

donde la restricción $U_n = \{(t, x) \in \mathbb{R} \times D \mid f(x^k) + (s^k)^t \cdot (x - x^k) \leq t, k = 0, \dots, n\}$. De la proposición 2.2 y la definición de subdiferencial tenemos que los ψ_n son funciones convexas para cada $n \in \mathbb{N}$ y además son lineales por partes. De la definición de subdiferencial se tiene

$$f(x) \geq \psi_{n+1}(x) \geq \psi_n(x), \quad \forall x \in \mathbb{R}^n.$$

Observe que a medida que n crece, la aproximación de ψ_n a la función f es cada vez mejor, aunque computacionalmente sea lo contrario. Podemos esperar entonces, que

las soluciones de los subproblemas (C_n) se aproximen cada vez más a las soluciones del problema (C).

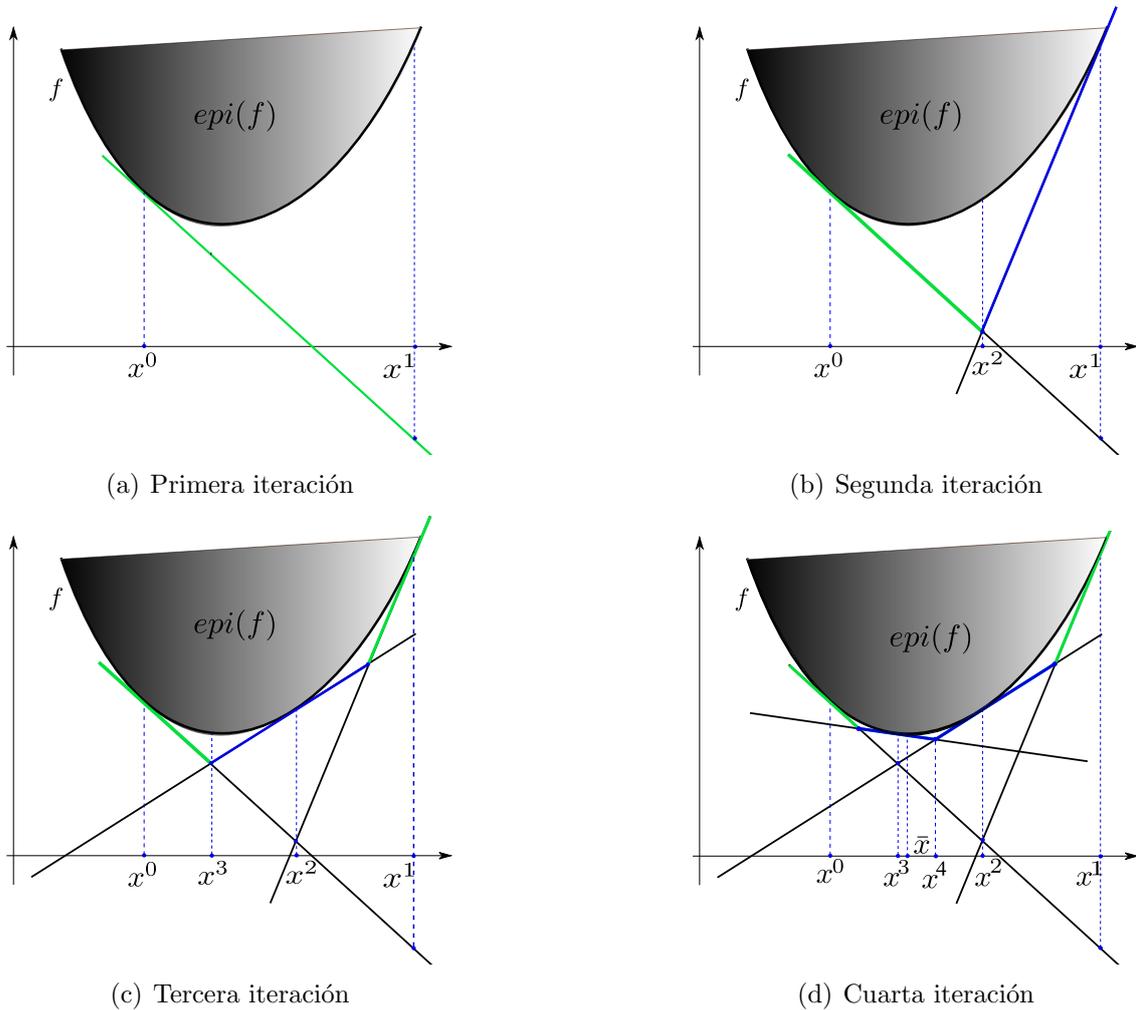


Figura 4.3: Cuatro primeras iteraciones del método de Planos Cortantes

A continuación se muestra el algoritmo generado por este método.

Algoritmo 4.2.1 (Algoritmo de Planos Cortantes). .

Paso 1. *Escoja cualquier punto $x^0 \in D$, y tome $k = 0$*

Paso 2. *Halle los valores de $f(x^k)$ y $s^k \in \partial f(x^k)$*

Paso 3. *Determine la iteración x^{k+1} el cual es una solución del problema (C_k)*

Paso 4. *Tome como siguiente iteración $k = k + 1$, e ir al paso 2.*

Teorema 4.4 (Convergencia del método Planos Cortantes). *Sea $D \subseteq \mathbb{R}^2$ un conjunto convexo compacto y $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa en \mathbb{R}^n . Entonces, para toda*

secuencia $\{x^k\}$ generada por el Algoritmo 4.2.1, se tiene que

$$\lim_{k \rightarrow \infty} \psi_k(x^{k+1}) = \bar{v} = \liminf_{k \rightarrow \infty} f(x^k),$$

donde $\bar{v} = \min_{x \in D} f(x)$, y las funciones $\psi_k : \mathbb{R}^n \rightarrow \mathbb{R}$ son definidas por $\psi_k(x) = \max_{j=1, \dots, k} \{f(x^j) + (s^j)^t \cdot (x - x^j)\}$.

4.2.1. Ventajas y desventajas del método de Planos Cortantes

Ventajas

Una ventaja importante de este método en comparación del método del Subgradiente es la posibilidad de construir una regla de parada informativa, implementable a través del decrecimiento nominal, es decir

$$\Delta^k = f(x^{k+1}) - \psi_{k+1}(x^{k+1}) \geq 0$$

y el algoritmo terminará cuando Δ^k se hace pequeño.

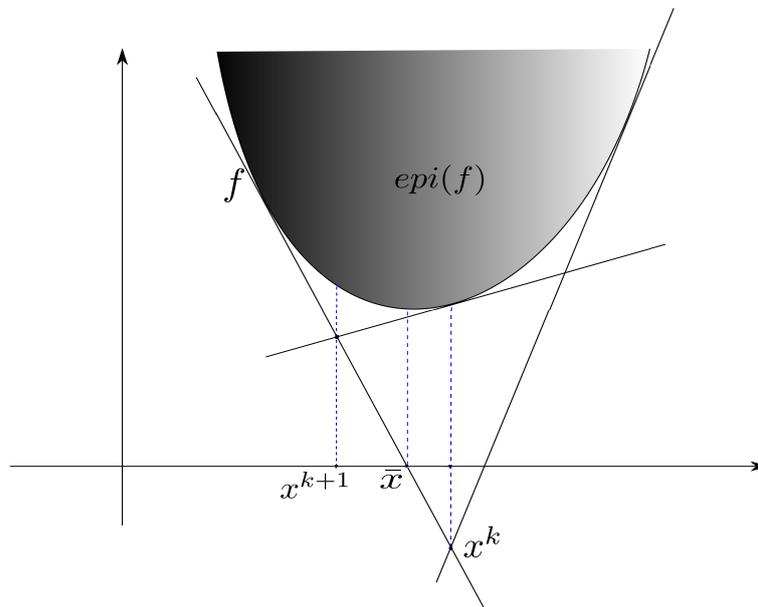


Figura 4.4: El gráfico muestra un crecimiento de f en la iteración x^{k+1} , pero ella se aleja de la solución óptima

Desventajas

Debido a que el método de Planos Cortantes se basa en la acumulación de subproblemas, y por tanto en la resolución de estos subproblemas cada vez mayores, ellos ocuparán mucha memoria en el ordenador y por lo tanto genera errores de redondeo computacional. Otra desventaja es la inestabilidad del método, ya que, al igual que el método de Subgradiente, ella no asegura el decrecimiento de la función objetivo en cada iteración. Este hecho puede ser observado en la figura 4.4 note que el punto x^k se encuentra $\Delta_k > \text{tol}$, donde la tol es una tolerancia de parada, y en la siguiente iteración se tiene que $f(x^{k+1}) > f(x^k)$ y además la iteración se aleja del valor óptimo. Es por esta razón que decimos que el algoritmo no está libre de inestabilidad. Para detalles sobre la convergencia del algoritmo Planos Cortantes, vea por ejemplo [2] y [28].

En la siguiente sección veremos el método de Bundle el cual se basa en la idea de aproximaciones cuasilineales de la función objetivo, como el método de Planos Cortantes complementando con un mecanismo de estabilización, con una regla de parada confiable, y una técnica que permite total control del tamaño de subproblemas lineales que se resuelven. Este método puede ser encontrado para más detalles vea por ejemplo [2],[28] y [29].

4.3. Método de Bundle

Considere el problema de optimización sin restricciones

$$(F) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s.a} & x \in \mathbb{R}^n \end{cases}$$

donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa en \mathbb{R}^n , no necesariamente diferenciable.

Supongamos que tenemos $x^k \in \mathbb{R}^n$, una aproximación de la solución del problema (F) (tal que hace decrecer a la función objetivo), de manera que la siguiente iteración $k+1$ tenga acumulada los puntos $z^i \in \mathbb{R}^n$ y los subgradietes $y^i \in \partial f(z^i)$, $i = 1, \dots, k$. El conjunto de información, de las secuencias $\{z^i\}_{i \in B}$ y $\{y^i\}_{i \in B}$, donde $B = \{1, 2, \dots, k\}$, es llamado el bundle. Entonces, la siguiente iteración z^{k+1} se calcula resolviendo el

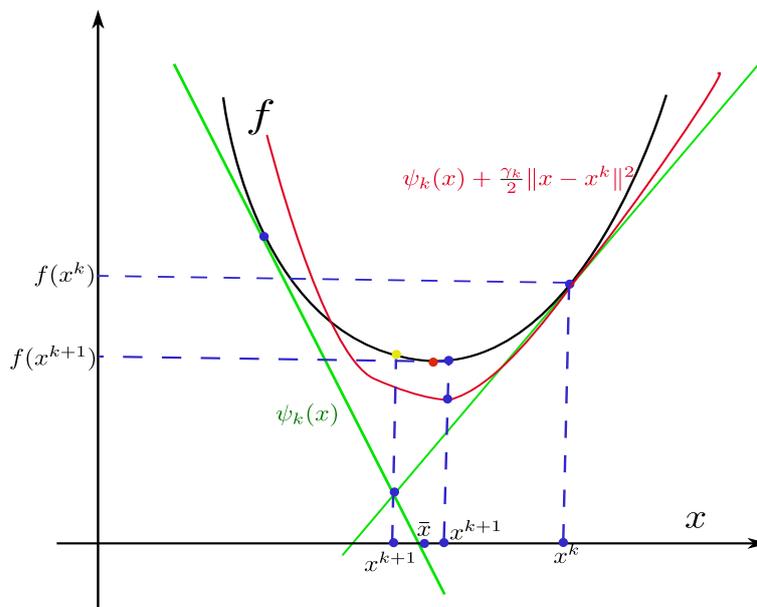


Figura 4.5: Un paso factible del método de Bundle genera en la iteración x^{k+1} un decrecimiento de f , es decir, $f(x^{k+1}) < f(x^k)$. Además podemos observar del gráfico que x^{k+1} es una buena aproximación a la solución \bar{x} con respecto a la aproximación del método de Planos Cortantes.

siguiente problema:

$$\begin{aligned} \text{minimizar} \quad & \psi_k(x) + \frac{\gamma_k}{2} \|x - x^k\|^2 \\ \text{s.a} \quad & x \in \mathbb{R}^n \end{aligned} \tag{4.6}$$

donde

$$\psi_k : \mathbb{R}^n \rightarrow \mathbb{R}. \quad \psi_k(x) = \max_{i=0,1,\dots,k} \{f(z^i) + \langle y^i, x - z^i \rangle\} \tag{4.7}$$

es la aproximación de Planos Cortantes de la función objetivo f y $\gamma_k > 0$ es llamado parámetro estabilizador. Este término estabilizador junto con el término cuadrático hacen el efecto de la iteración $k + 1$ (z^{k+1}) respecto a la iteración k (x^k) sea la más cercana posible, donde x^k es llamado centro de estabilización.

El método de Bundle que presentamos, tiene como siguiente iteración a aquellos puntos donde el centro de estabilización cambia, es decir, cuando el paso x^k a z^{k+1} efectúa un decrecimiento en la función objetivo (esto es, que sea una dirección de descenso). Si sucediera eso, entonces tomamos como punto $x^{k+1} = z^{k+1}$. Este tipo de paso es llamado paso factible. En caso contrario, el centro estabilizador no cambia, es decir, tomamos $x^{k+1} = x^k$ y agregamos un nuevo plano cortante. Este tipo de iteración es llamada paso

no factible.

Método de Bundle proximal

Antes de la implementación del algoritmo determinaremos una forma de solucionar el problema (4.6), el cual tiene solución única debido a que la función objetivo es fuertemente convexa, ver Teorema 2.13, y sacaremos información del bundle acumulado de la ecuación (4.7), pero sin perder en el proceso iterativo de la convergencia a la solución del problema.

Sea

$$\psi_k(x) = f(x^k) + \max_{i=0,1,\dots,k} \{-e_i^k + (y^i)^t \cdot (x - x^k)\} \quad (4.8)$$

donde

$$e_i^k = f(x^k) - f(z^i) - (y^i)^t \cdot (x^k - z^i) \geq 0, \quad i = 0, 1, \dots, k \quad (4.9)$$

entonces por la definición de ϵ -subdiferencial se tiene:

$$y^i \in \partial_{e_i^k} f(x^k), \quad i = 0, 1, \dots, k \quad (4.10)$$

Ahora, supongamos que hemos podido eliminar algunos elementos del bundle, es decir, algunos planos cortantes, entonces

$$\psi_k(x) = f(x^k) + \max_{i \in B_k} \{-e_i^k + (y^i)^t \cdot (x - x^k)\} \quad (4.11)$$

donde B_k es el conjunto de índices que define el bundle actual, y se tendrá:

$$y^i \in \partial_{e_i^k} f(x^k), \quad i \in B_k \quad (4.12)$$

Para el proceso de eliminación de algunos elementos del bundle y para no perder la convergencia se debe definir un nuevo plano cortante que sustituya a los que ya no se encuentre en el bundle. Esta función es llamada función agregada, para tal efecto,

primero se redefine (4.6) con restricción (4.11) al problema equivalente

$$\begin{aligned} \text{minimizar} \quad & t + \frac{\gamma^k}{2} \|x - x^k\|^2 \\ \text{s.a} \quad & (t, x) \in U \end{aligned} \quad (4.13)$$

$$U = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n \mid f(x^k) - e_i^k + (y^i)^t(x - x^k) \leq t, i \in B_k\} \quad (4.14)$$

Para obtener la solución de este problema cuadrático, procedemos a resolver en su forma dual, el cual es dado por:

$$\text{minimizar} \quad \frac{1}{2} \left\| \sum_{i \in B_k} \nu_i y^i \right\|^2 + \gamma_k \sum_{i \in B_k} \nu_i e_i^k \quad (4.15)$$

$$\nu \in \left\{ \nu \in \mathbb{R}_+^{|B_k|} \mid \sum_{i \in B_k} \nu_i = 1 \right\} \quad (4.16)$$

Para continuar con la idea de construcción de la función agregada, veamos el siguiente resultado que determina una secuencia que converge a una solución del problema (4.6).

Lema 4.1 (Ver demostración [2]). *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa en \mathbb{R}^n . Para cualquier $x^k \in \mathbb{R}^n$, $y^i \in \partial_{e_i^k} f(x^k)$, $i \in B_k$, y $\gamma_k > 0$, definimos*

$$d^k = \sum_{i \in B_k} \nu_i^k y_i^k \quad (4.17)$$

donde $\nu^k \in \mathbb{R}^{|B_k|}$ es una solución de (4.15)-(4.16). Entonces, la única solución z^{k+1} del problema (4.6) con función objetivo definida por (4.11), es dada por

$$z^{k+1} = x^k - \frac{1}{\gamma_k} d^k \quad (4.18)$$

Además, tenemos que

$$d^k \in \partial_{\epsilon_k} f(x^k) \quad (4.19)$$

donde

$$\epsilon_k = \sum_{i \in B_k} \nu_i^k e_i^k \geq 0. \quad (4.20)$$

Con este resultado estamos en la facultad de definir lo que habíamos llamado función agregada, el cual es dada por

$$\psi_k^a : \mathbb{R}^n \rightarrow \mathbb{R}, \psi_k^a(x) = f(x^k) - \epsilon_k + (d^k)^t \cdot (x - x^k) \quad (4.21)$$

Entonces, se ha obtenido una técnica de cómo retirar planos cortantes sin perder convergencia, el cual es agregando la función agreda ψ_k^a , y así disminuyendo el bundle que computacionalmente es estable.

Seguidamente, para generar el algoritmo, definimos la expresión:

$$e_{k+1}^{k+1} = f(x^k) - f(z^{k+1}) - (y^{k+1})^t \cdot (x^k - z^{k+1}) \geq 0 \quad (4.22)$$

Además, debemos garantizar la validez de la ecuación (4.12), con $k + 1$ en vez de k . Como es fácil verificar, la fórmula siguiente garantiza tal condición:

$$e_i^{k+1} = e_i^k + f(x^{k+1}) - f(x^k) + (y^i)^t \cdot (x^k - x^{k+1}), \forall i \in B_{k+1}/\{k+1\} \quad (4.23)$$

Para más detalles de este proceso, ver [2]. Realizado estos pasos, presentamos el algoritmo de Bundle para una programación cuadrática.

Algoritmo 4.3.1 (El algoritmo Bundle).

Fijar un número entero $M \geq 2$ (el tamaño máximo permitido del bundle). Escoger un $x^0 \in \mathbb{R}^n$ y tomar $k = 0$, $\mathcal{K} = \emptyset$. Escoger $\sigma \in \langle 0, 1 \rangle$. Calcule $f(x^0)$ e $y^0 \in \partial f(x^0)$, tomar $z^0 = x^0$, $e_0^0 = 0$, y $B_0 = \{0\}$ y definir el bundle inicial que contiene el único par (y^0, e_0^0) .

Paso 1. *Escoger $\gamma_k > 0$ y calcular z^{k+1} , la solución del problema (4.6), con función objetivo dado por (4.11).*

Paso 2. *Halle $f(x^{k+1})$, $y^{k+1} \in \partial f(z^{k+1})$ y*

$$\Delta_k = f(x^k) - \psi_k(z^{k+1}) - \frac{\gamma_k}{2} \|z^{k+1} - x^k\|^2$$

Paso 3. *Si $f(x^k) - f(z^{k+1}) \geq \sigma \Delta_k$*

Tome $x^{k+1} = z^{k+1}$ y concluir que la iteración k se encuentra en el conjunto \mathcal{K} (un paso factible)

Caso contrario, tome $x^{k+1} = x^k$ (un paso no factible)

Paso 4. Si $|B_k| < M$, tome $B_{k+1} = B_k \cup \{k+1\}$.

Si $|B_k| = M$, escoger los índices $i_1, i_2 \in B_k$ y tome $B_{k+1} = (B_k - \{i_2\}) \cup \{k+1\}$,
 $y^{i_1} = d^k$, $e_{i_1}^k = \epsilon_k$ donde $d^k \in \mathbb{R}^n$, y $\epsilon_k \geq 0$ son definidos por el lema anterior.

Paso 5. Si $k \in \mathcal{K}$, calcular e_i^{k+1} , $i \in B_{k+1} - \{k+1\}$, por la fórmula (4.23).

Caso contrario, tome $e_i^{k+1} = e_i^k$, $i \in B_{k+1} - \{k+1\}$. Calcule e_{k+1}^{k+1} por la fórmula dada en (4.22). Define el nuevo bundle como el conjunto de pares (y^i, e_i^{k+1}) , $i \in B_{k+1}$.

Paso 6. Tome $k = k+1$ y retorne al **paso 1**.

Teorema 4.5 (Convergencia de los pasos factibles del método Bundle). Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa en \mathbb{R}^n . Supongamos que el algoritmo 4.3.1 genera una secuencia de iteraciones para las cuales el conjunto \mathcal{K} de pasos factibles es infinito. Supongamos que la secuencia $\{\gamma_k\}$ satisfaga la condición

$$\sum_{k \in \mathcal{K}} \frac{1}{\gamma_k} = +\infty.$$

Para la secuencia $\{x^k\}$ del algoritmo 4.3.1 se tiene que

$$\lim_{k \rightarrow \infty} f(x^k) = \bar{v},$$

donde $\bar{v} = \inf_{x \in \mathbb{R}^n} f(x)$, y para $k \in \mathcal{K}$, se verifica

$$\Delta_k \rightarrow 0 \quad (k \rightarrow \infty).$$

Si el problema (F) tiene una solución y

$$\gamma_k \geq \tilde{\gamma} > 0 \quad \forall k \in \mathcal{K},$$

donde el número $\tilde{\gamma}$ no depende de k , entonces $\{x^k\}$ converge a una solución del problema (F).

Teorema 4.6 (Convergencia de pasos no factibles del método Bundle). Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa en \mathbb{R} . Supongamos que el algoritmo 4.3.1 genera una secuencia $\{x^k\}$ para la cual el conjunto de índice de paso factibles \mathcal{K} es finito, es decir, $x^k =$

$x^{\bar{k}} \forall k \geq \bar{k}$ para algún índice \bar{k} . Supongamos que la secuencia $\{\gamma_k\}$ sea limitada y que ella satisface la condición

$$\gamma_{k+1} \geq \gamma_k \forall k \geq \bar{k}.$$

Entonces, $x^{\bar{k}}$ es una solución del problema (F), la secuencia $\{z^k\}$ converge a $x^{\bar{k}}$, y $\Delta_k \rightarrow 0$ ($k \rightarrow \infty$).

Para ver más detalles de los métodos de Bundle y de la convergencia del algoritmo 4.3.1, y el bundle generado en esta sección, vea por ejemplo [2],[28] y [29].

Capítulo 5

Método de Direcciones Factibles para Optimización Convexa no diferenciable

En este capítulo se presentará un método de punto interiores de un problema convexo sin restricciones no necesariamente diferenciable, conocido como NFDA (A feasible directions method for nonsmooth convex optimization) que fue realizado por Wilhelm Passarella, Mario Tanaka Fo, Alfredo Canales y José Herskovits. Este método, que determina la minimización de un problema de optimización convexa no diferenciable sin restricciones, resuelve el problema llevándolo a su forma equivalente de optimización lineal con restricciones de desigualdad no diferenciable, para luego hacer una aproximación a las restricciones y así obtener un problema lineal con restricciones de desigualdad diferenciable, como el método de Planos Cortantes, pero sin la formación de subproblemas. En cada iteración, el algoritmo NFDA encuentra una dirección de búsqueda factible y de descenso hallados en dos sistemas lineales obtenido por el método FDI-PA, tomando como punto inicial en el interior del $\text{epi}(f)$. Para obtener la dirección de búsqueda factible y de descenso, este método usa la idea de punto factible y no factible realizada por el método Bundle que fue visto en el capítulo anterior. A continuación es presentado el método NDFA.

5.1. El Método de Direcciones Factibles no Diferenciables

Consideremos el problema de optimización convexa sin restricciones

$$(P) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s.a} & x \in \mathbb{R}^n \end{cases}$$

Donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa no necesariamente diferenciable. Para la existencia de la solución del problema y la convergencia del algoritmo que se genera por el método NFDA consideramos las siguientes hipótesis:

Hipótesis 4.1: Existe para f un conjunto de nivel acotado.

Hipótesis 4.2: Para cada $a \in \mathbb{R}$ y cada punto $x \in \Gamma_a = \{x \in \mathbb{R}^n \mid f(x) \leq a\}$, podemos determinar un $s \in \partial f(x)$.

5.1.1. Descripción del Método NFDA

Se reformula el problema primal (P) a su forma lineal equivalente

$$(P^*) \quad \begin{cases} \text{minimizar} & \varphi(x, z) = z \\ \text{s.a} & (x, z) \in \mathbb{R}^{n+1} \\ & f(x) \leq z \end{cases}$$

el cual es un problema de optimización lineal diferenciable con restricción de desigualdad no necesariamente diferenciable, con $\varphi : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ una función lineal diferenciable. Empleando la idea del método de Plano Cortantes, se hace una aproximación a la restricción del problema (P^*) , y suponiendo que estamos en la iteración k , y que el conjunto actual de planos cortantes es dado

$$g_i^k(x, z) = f(y_i^k) + (s_i^k)^t(x - y_i^k) - z, \quad i = 0, 1, \dots, l$$

donde $y_i^k \in \mathbb{R}^n$ son puntos auxiliares, $s_i^k \in \partial f(y_i^k)$ son los subgradientes en el bundle encontrado. Definiendo

$$\tilde{g}_l^k(x, z) = [g_0^k(x, z), \dots, g_l^k(x, z)]^t, \quad \tilde{g}_l^k : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^{l+1},$$

de tal proceso se tiene una aproximación (AP_l) del problema equivalente (P^*) , el cual es un problema lineal con restricciones de desigualdad diferenciable, es decir,

$$(AP_l) \quad \begin{cases} \text{minimizar} & \varphi(x, z) = z \\ & (x, z) \in \mathbb{R}^{n+1} \\ \text{s.a} & \tilde{g}_l^k(x, z) \leq 0 \end{cases}$$

Es cierto que el problema de aproximación (AP_l) puede que no alcance un mínimo, pero el método no busca resolver este problema, solo busca una dirección de descenso d_l^k factible, en el interior del conjunto de restricciones del problema de aproximación, el cual es encontrado por el método FDIPA que se desarrollamos en el capítulo 3.

Respecto a la longitud de paso se toma

$$t = \max\{t \mid \tilde{g}_l^k((x^k, z^k) + td_l^k) \leq 0\}$$

pero desde que t no es siempre finito, se tomará como longitud de paso

$$t_l^k = \min\{t_{max}/\mu, t\}$$

donde $\mu \in \langle 0, 1 \rangle$ y $t_{max} > 0$. Entonces el siguiente punto interior será tomado como

$$(x_{l+1}^k, z_{l+1}^k) = (x^k, z^k) + t_l^k d_l^k \quad (5.1)$$

el cual es factible con respecto al problema (AP_l) . Veamos ahora que sea factible para el problema (P^*) , para ello, tomando como punto auxiliar

$$(y_{l+1}^k, w_{l+1}^k) = (x^k, z^k) + \mu t_l^k d_l^k \quad (5.2)$$

El punto (y_{l+1}^k, w_{l+1}^k) será estrictamente factible, en el sentido que se encuentra en el interior del epígrafo de f , respecto al problema (P) , que es $w_{l+1}^k > f(y_{l+1}^k)$ y se considera que el conjunto de planos cortantes es una buena aproximación para f en una vecindad de x^k . Si el punto auxiliar $w_{l+1}^k > f(x_{l+1}^k)$ es satisfecha, entonces diremos que el paso es factible y se toma como un nuevo punto $x^{k+1} = y_{l+1}^k$, $z^{k+1} = w_{l+1}^k$. Caso contrario,

diremos que el paso es no factible y realizamos una nueva aproximación agregando un nuevo plano cortante $g_{l+1}^k(x, z)$ al bundle actual.

Realizado la explicación del método NFDA, este genera el siguiente algoritmo llamado algoritmo **NFDA**, el cual es descrito seguidamente.

5.2. Algoritmo de Direcciones Factibles no Diferenciables - NFDA

Paso 1. Calcule la subgradiente $s_l^k \in \partial f(y_l^k)$, y halle un nuevo plano cortante, para la iteración (x^k, z^k) , que es definido por

$$g_l^k(x, z) = f(y_l^k) + (s_l^k)^t(x - y_l^k) - z$$

Considerando $\nabla g_l^k(x, z) = \begin{bmatrix} s_l^k \\ -1 \end{bmatrix} \in \mathbb{R}^{n+1}$ para el conjunto de planos cortantes y definiendo

$$\tilde{g}_l^k(x, z) = [g_0^k(x, z), \dots, g_l^k(x, z)]^t \in \mathbb{R}^{l+1}$$

cuya Jacobiana es dada por

$$J\tilde{g}_l^k(x, z) = [\nabla g_0^k(x, z), \dots, \nabla g_l^k(x, z)]^t \in \mathbb{R}^{(n+1) \times (l+1)}$$

Paso 2. Buscando la dirección factible d_l^k para el problema (AP_l) usando el método de FDIPA

(ia) Calcule $d_{\alpha l}^k$ y $\lambda_{\alpha l}^k$, resolviendo

$$B^k d_{\alpha l}^k + J\tilde{g}_l^k(x^k, z^k)\tilde{\lambda}_{\alpha l}^k = -\nabla\psi(x^k, z^k) \quad (5.3)$$

$$\tilde{\Lambda}_l^k [J\tilde{g}_l^k(x^k, z^k)]^t d_{\alpha l}^k + \tilde{G}_l^k(x^k, z^k)\tilde{\lambda}_{\alpha l}^k = 0 \quad (5.4)$$

(ib) Calcule $d_{\beta l}^k$ y $\lambda_{\beta l}^k$, resolviendo

$$B^k d_{\beta l}^k + J\tilde{g}_l^k(x^k, z^k)\tilde{\lambda}_{\beta l}^k = 0 \quad (5.5)$$

$$\tilde{\Lambda}_l^k [J\tilde{g}_l^k(x^k, z^k)]^t d_{\beta l}^k + \tilde{G}_l^k(x^k, z^k)\tilde{\lambda}_{\beta l}^k = -\tilde{\lambda}_l^k \quad (5.6)$$

donde $\tilde{\lambda}_{\alpha l}^k = (\lambda_{\alpha 0}^k, \dots, \lambda_{\alpha l}^k)$, $\lambda_{\beta l}^k = (\lambda_{\beta 0}^k, \dots, \lambda_{\beta l}^k)$, $\tilde{\lambda}_l^k = (\tilde{\lambda}_0^k, \dots, \tilde{\lambda}_l^k)$,

$\tilde{\Lambda}_l^k = \mathbf{diag}(\lambda_0^k, \dots, \lambda_l^k)$ y $\tilde{G}_l^k(x, z) = \mathbf{diag}(g_0^k(x, z), \dots, g_l^k(x, z))$.

(ii) Si $(d_{\beta l}^k)^t \nabla \psi(x^k, z^k) > 0$,

$$\rho = \min \left\{ \varphi \|d_{\alpha l}^k\|^2, (\xi - 1) \frac{(d_{\alpha l}^k)^t \nabla \psi(x^k, z^k)}{(d_{\beta l}^k)^t \nabla \psi(x^k, z^k)} \right\}$$

Caso contrario se toma

$$\rho = \varphi \|d_{\alpha l}^k\|^2$$

(iii) Calculando la dirección factible

$$d_l^k = d_{\alpha l}^k + \rho d_{\beta l}^k.$$

Paso 3. Calculando la longitud de paso

$$t_l^k = \min \{ t_{max}/\mu, \max\{t \mid \tilde{g}_l^k((x^k, z^k) + td_l^k) \leq 0\} \} \quad (5.7)$$

Paso 4. Hallando el nuevo punto interior

i) Define el punto auxiliar $(y_{l+1}^k, w_{l+1}^k) = (x^k, z^k) + \mu t_l^k d_l^k$,

ii) Si $w_{l+1}^k \leq f(y_{l+1}^k)$ se tendrá un paso no factible y en este caso definimos

un

nuevo $\tilde{\lambda}_{l+1}^k > 0$, y_{l+1}^k (nuevo punto auxiliar) y $l = l + 1$.

Caso contrario, se tendrá un paso factible y tomamos $(x^{k+1}, z^{k+1}) = (y_{l+1}^k, w_{l+1}^k)$,

definimos $\tilde{\lambda}_0^{k+1} > 0$, B^{k+1} simétrica y definida positiva y tomar $k = k + 1$,

e iniciamos un nuevo conjunto de planos cortantes $l = 0$, $y_0^k = x^k$.

iii) Ir al paso 1.

Para tomar los valores de actualización de λ y B se debe tener en cuenta la satisfacción de las siguientes hipótesis:

Hipótesis 4.3: Los gradientes de las direcciones activas en (x^*, z^*) son linealmente independientes, es decir, los vectores $\nabla g_i(x^*, z^*)$ son linealmente independientes, para cada i tal que $g_i(x^*, z^*) = 0$.

Hipótesis 4.4: Existen $\lambda^I, \lambda^S > 0$ tales que $\lambda^I \leq \lambda_i^k \leq \lambda^S, \forall i = 1, \dots, l, \forall k \in \mathbb{N}$.

Hipótesis 4.5 Existen números positivos σ_1 y σ_2 tales que

$$\sigma_1 \|\nu\|^2 \leq \nu^T B \nu \leq \sigma_2 \|\nu\|^2$$

para cualquier $\nu \in \mathbb{R}^{n+1}$.

Hipótesis 4.6 Existen $\omega_1, \omega_2 > 0$ tales que $\omega_1 \leq \omega_i^k \leq \omega_2, \forall k \in \mathbb{N}$ y $i = 1, \dots, m$.

5.3. Análisis de convergencia del algoritmo NFDA

En esta sección, se demostrará la convergencia global del algoritmo NFDA. Para ello, primero se mostrará que la dirección d_l^k es una dirección de descenso para la función ψ del problema auxiliar. Mostraremos que el número de pasos no factibles en cada iteración es finito, es decir, desde que $(x^k, z^k) \in \text{int}(\text{epi}(f))$, después de un número finito de sub-iteraciones, se obtiene un punto $(x^{k+1}, z^{k+1}) \in \text{int}(\text{epi}(f))$. En consecuencia, la sucesión $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ generada por el algoritmo es acotada y pertenece al interior del epígrafo de f . Se demuestra que el algoritmo generado es determinado por un criterio de parada, esto es, mostrando que cuando $k \rightarrow \infty$ la dirección d^k converge a cero. Finalmente se demuestra que cualquier punto de acumulación (x^*, z^*) de la secuencia $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ generada por el algoritmo NFDA satisfacen la condición de optimalidad $0 \in \partial f(x^*)$ y, por lo tanto, por el teorema 2.15, el punto de acumulación de la secuencia $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ será una solución del problema (P) .

Lema 5.1. Sean el conjunto de vectores $\{\nabla g_i(x, z) \mid g_i(x, z) = 0\}_l$ un conjunto linealmente independiente, cualquier $\bar{\lambda} \in \mathbb{R}^{n+1}$ positivo, cualquier vector $(x, z) \in \text{int}(\text{epi} f)$ y

cualquier matriz simétrica definida positiva $B \in \mathbb{R}^{(n+1) \times (n+1)}$, la matriz

$$\begin{bmatrix} B & J\tilde{g}(x, z) \\ \tilde{\Lambda}[J\tilde{g}(x, z)]^t & \tilde{G}(x, z) \end{bmatrix}$$

es no singular.

Prueba: Para demostrar la no singularidad debemos de verificar que el sistema

$$\begin{aligned} Bd + J\tilde{g}(x, z)\bar{\lambda} &= 0 \\ \tilde{\Lambda}[J\tilde{g}(x, z)]^t d + \tilde{G}\bar{\lambda} &= 0 \end{aligned}$$

tiene como solución única la trivial. En efecto,

$$Bd + \sum_{i=1}^l \bar{\lambda}_i \nabla \tilde{g}_i(x, z) = 0 \quad (5.8)$$

$$\lambda_i [\nabla g_i^t(x, z)]^t \cdot d + g_i(x, z) \bar{\lambda}_i = 0, \quad i = 1, \dots, l \quad (5.9)$$

Sacando transpuesta y multiplicando por d a la ecuación (5.8), se obtiene

$$d^t Bd + \sum_{i=1}^l \bar{\lambda}_i [\nabla \tilde{g}_i^t]^t \cdot (x, z) d = 0$$

de la ecuación (5.9), $[\nabla g_i(x, z)]^t \cdot d = -\frac{\bar{\lambda}_i}{\lambda_i} g_i(x, z)$, y reemplazando en la expresión anterior

$$d^t Bd + \sum_{i=1}^l \frac{\bar{\lambda}_i^2}{\lambda_i} (-g_i(x, z)) = 0,$$

y desde que B es definida positiva, $g(x, z) < 0$, ya que $(x, z) \in \text{epi}(f)$, y $\lambda_i > 0$, para $i = 1, \dots, l$, se tiene que $d = 0$ y $\bar{\lambda} = 0$. \square

El siguiente lema muestra que d_α es una dirección de descenso para la función ψ del problema de aproximación (AP_l) .

Lema 5.2. *El vector d_α satisface*

$$d_\alpha^t \cdot \nabla \psi(x, z) \leq -d_\alpha^t B d_\alpha$$

Prueba: Se sigue de la ecuación (5.3)

$$d_\alpha^t B d_\alpha + d_\alpha^t J \tilde{g}(x, z) \lambda_\alpha = -d_\alpha^t \nabla \cdot \psi(x, z) \quad (5.10)$$

y de la ecuación (5.4)

$$d_\alpha^t J \tilde{g}(x, z) = -\lambda_\alpha^t \tilde{\Lambda}^{-1} \tilde{G}(x, z) \quad (5.11)$$

Sustituyendo (5.11) en (5.10) tenemos

$$d_\alpha^t \cdot \nabla \psi(x, z) = -d_\alpha^t B d_\alpha + \lambda_\alpha^t \tilde{\Lambda}^{-1} \tilde{G}(x, z) \lambda_\alpha$$

Desde que $\tilde{\Lambda}^{-1}$ es definida positiva y por la hipótesis 4.4, 4.5 y del hecho que $\tilde{\Lambda}^{-1} \tilde{G}(x, z)$ es semidefinida negativa, se tiene

$$d_\alpha^t \cdot \nabla \psi(x, z) \leq -d_\alpha^t B d_\alpha$$

□

Proposición 5.1. *La dirección $d = d_\alpha + \rho d_\beta$ es una dirección de descenso para la función objetivo del problema (AP_l) en un punto (x, z) .*

Prueba: Como $d = d_\alpha + \rho d_\beta$, se tiene

$$d^t \cdot \nabla \psi(x, z) = d_\alpha^t \cdot \nabla \psi(x, z) + \rho d_\beta^t \cdot \nabla \psi(x, z)$$

en el caso cuando $d_\beta^t \cdot \nabla \psi(x, z) > 0$, tenemos

$$\rho \leq (\xi - 1) \frac{d_\alpha^t \cdot \nabla \psi(x, z)}{d_\beta^t \cdot \nabla \psi(x, z)}.$$

Por lo tanto $d^t \nabla \cdot \psi(x, z) \leq d_\alpha^t \cdot \nabla \psi(x, z) + (\xi - 1) d_\alpha^t \nabla \cdot \psi(x, z) = \xi d_\alpha^t \cdot \nabla \psi(x, z) < 0$.

Por otro lado, cuando $d_\beta^t \cdot \nabla \psi(x, z) \leq 0$, se tiene por el lema 5.2 que $d^t \cdot \nabla \psi(x, z) \leq d_\alpha^t \cdot \nabla \psi(x, z) < 0$ para cualquier $\rho > 0$. □

Como consecuencia de la proposición anterior tenemos que del punto $(y^{k+1}, z^{k+1}) = (x^k, z^k) + \mu t_l (d_x^l, d_z^l)$, con $\nabla \psi(x, z) = (0, 0, \dots, 1)^t$ la secuencia $\{z^k\}_{k \in \mathbb{N}}$ es decreciente,

es decir, $z^{k+1} < z^k$ para cualquier k . Además, la secuencia $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ generada por el presente algoritmo es acotado, desde que ella pertenece al conjunto

$$\text{int}(\text{epi}f) \cap \{(x, z) \in \mathbb{R}^{n+1} \mid z \leq z^0\}$$

y el conjunto de nivel de f es acotado y no vacío.

Lema 5.3. *Sea $X \subseteq \mathbb{R}^n$ un conjunto convexo. Considere $x_0 \in \text{int}(X)$ y $\bar{x} \in \overline{X}$. Sea $\{\bar{x}^k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n - X$ una secuencia tal que $\bar{x}^k \rightarrow \bar{x}$. Sea $\{x^k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ una secuencia definida por $x^k = x_0 + \mu(\bar{x}^k - x_0)$ con $\mu \in (0, 1)$. Entonces $\exists k_0 \in \mathbb{N}$ tal que $x^k \in \text{int}(X)$, $\forall k > k_0$.*

Prueba: De la secuencia

$$x^k = x_0 + \mu(\bar{x}^k - x_0) \rightarrow x_0 + \mu(\bar{x} - x_0) = x_\mu.$$

Desde que el segmento $[\bar{x}, x_0] \subseteq X \cup \partial X$ y $\mu \in (0, 1)$ se tiene que $x_\mu \in \text{int}(X)$ y en consecuencia existe $\delta > 0$ tal que $B(x_\mu, \delta) \subset \text{int}(X)$. Desde que $x^k \rightarrow x_\mu$, entonces $\exists k_0 \in \mathbb{N}$ tal que $x^k \in B(x_\mu, \delta) \subseteq \text{int}(X)$, $\forall k > k_0$. \square

Proposición 5.2. *Sea $(\tilde{x}^k, \tilde{z}^k)$ un punto de acumulación de la secuencia $\{(x_l^k, z_l^k)\}_{l \in \mathbb{N}}$ generada por el algoritmo NFDA, y definida en la ecuación (5.1) para k fijo, entonces $\tilde{z}^k = f(\tilde{x}^k)$.*

Prueba: De la secuencia $\{(x_l^k, z_l^k)\}_{l \in \mathbb{N}}$ definida por el problema $\{(AP_l)\}$ se tiene que $\tilde{z}^k \leq f(\tilde{x}^k)$. Suponiendo que tenemos $\tilde{z}^k < f(\tilde{x}^k)$. Desde que $\{(\tilde{x}^k, \tilde{z}^k)\}$ es un punto de acumulación de la secuencia $\{(x_l^k, z_l^k)\}_{l \in \mathbb{N}}$, entonces existe $\mathbb{N}' \subseteq \mathbb{N}$ tal que $\{(x_l^k, z_l^k)\}_{l \in \mathbb{N}'}$ converja a $\{(\tilde{x}^k, \tilde{z}^k)\}$ tal que $(s_l^k)_{l \in \mathbb{N}'} \rightarrow s^k$. Estas secuencias existen ya que $\{(x_l^k, z_l^k)\}_{l \in \mathbb{N}'}$ como $\{s_l^k\}_{l \in \mathbb{N}'}$ están en un conjunto compacto. Para cada $l \in \mathbb{N}'$ consideremos los correspondientes planos cortantes $f(x_l^k) + s_l^t \cdot (x - x_l^k) = z$, entonces para el punto \tilde{x}^k existirá un $z(\tilde{x}^k)$ tal que $f(x_l^k) + s_l^t(x - x_l^k) = z(\tilde{x}^k)$. Tomando límite para $l \rightarrow \infty$, tendremos $z(\tilde{x}^k) = f(\tilde{x}^k)$. Entonces, para $l \in \mathbb{N}'$ bien grande, tendremos que $(\tilde{x}^k, \tilde{z}^k)$ está debajo de los l planos cortantes, lo cual es una contradicción y por lo tanto $\tilde{z}^k = f(\tilde{x}^k)$. \square

Proposición 5.3. Sea $(x^k, z^k) \in \text{int}(\text{epi}f)$. La siguiente iteración $(x^{k+1}, z^{k+1}) \in \text{int}(\text{epi}f)$ es obtenida después de un número finito de sub-iteraciones.

Prueba: La prueba comienza observando el paso 4) en el algoritmo, el cual nos dice que $(x^{k+1}, z^{k+1}) = (y_{l+1}^k, w_{l+1}^k)$. Si $w_{l+1}^k > f(y_{l+1}^k)$, consecuentemente tenemos $(x^{k+1}, z^{k+1}) \in \text{int}(\text{epi}f)$. Caso contrario, como la secuencia $\{(x_l^k, z_l^k)\}_{l \in \mathbb{N}}$ es acotada por construcción, y por la proposición 5.2 se tiene que un punto de acumulación $(\tilde{x}^k, \tilde{z}^k)$ es tal que $\tilde{z}^k = f(\tilde{x}^k)$. Considere ahora las secuencias definidas por las ecuaciones (5.1) y (5.2),

$$(y_l^k, w_l^k) = (x^k, z^k) + \mu [(x_l^k, z_l^k) - (x^k, z^k)], \text{ para } \mu \in \langle 0, 1 \rangle$$

y del lema 5.3, existe $l_0 \in \mathbb{N}$ tal que $(y_l^k, w_l^k) \in \text{int}(\text{epi}f)$, para $l \geq l_0$. Pero esto es una condición para un paso factible y por lo tanto la prueba está completada. \square

Proposición 5.4. Sea (x^*, z^*) un punto de acumulación de la secuencia $\{(x^k, z^k)\}_{k \in \mathbb{N}}$. La secuencia de matrices $\{J\tilde{g}_l^k\}_{k, j \in \mathbb{N}}$ posee un punto de acumulación tal que por lo menos una de sus columnas es el gradiente de una restricción activa en (x^*, z^*) .

Prueba: Como la función f es convexa definida en todo \mathbb{R}^n tendremos $\partial f(x)$ es un conjunto acotado para cualquier $x \in \mathbb{R}^n$, y además como las secuencias $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ y $\{(x_l^k, z_l^k)\}_{k \in \mathbb{N}}$ son acotadas, tomando $l \in \mathbb{N}$ tal que $\{J\tilde{g}_l^k\}_{k \in \mathbb{N}}$ es acotada y por lo tanto posee un punto de acumulación, el cual lo denotamos por $J\tilde{g}_l^*$. Suponiendo que para cualquier l tomado, las columnas de $\nabla \tilde{g}_l^*$ estén formadas por gradientes de restricciones inactivas en (x^*, z^*) , entonces:

Desde que $g_r^k(x^*, z^*) = f(x^k) + \langle s^k, x^* - x^k \rangle - z^*$, donde $s^k \in \partial f(x^k)$, $1 \leq r \leq l$, de la condición que las restricciones son inactivas en (x^*, z^*) existe $\beta > 0$ tal que $g_r^k(x^*, z^*) < -\beta$, $\forall r \in \{1, \dots, l\}$ y tomando $k \rightarrow \infty$ obtenemos una contradicción. \square

De la proposición anterior podemos garantizar que el conjunto

$$I^* = \{i \in \{1, 2, \dots, l\} \mid g_i(x^*, z^*) = 0\}$$

es no vacío.

Lema 5.4. La matriz simétrica $U^k = (Jg_i^k)^t(B^k)^{-1}Jg_i^k - (\Lambda^k)^{-1}G^k$ de orden $l \times l$ es definida positiva $\forall k \in \mathbb{N}$ y existen $N \subseteq \mathbb{N}$ y $\eta > 0$ tales que $\|U^k\| > \eta, \forall k \in N$.

Prueba: Denotando $A^k = Jg_i^k$. Como $(B^k)^{-1}$ es definida positiva se sigue que $(A^k)^t(B^k)^{-1}A^k$ es definida positiva. Además, por construcción $(\Lambda^k)^{-1}G^k$ es semidefinida negativa. Luego U^k es semidefinida positiva. Tomando $v \in \mathbb{R}^l$ tal que $v^t U^k v = 0$. Entonces $\sum_{i=1}^l (\lambda_i^k)^{-1} g_i^k v_i^2 = v^t (A^k)^t (B^k)^{-1} A^k v \geq 0$. Pero como $(\lambda_i^k)^{-1} g_i^k < 0$ se tiene que $\sum_{i=1}^l (\lambda_i^k)^{-1} g_i^k v_i^2 \leq 0$. Luego, $v = 0$ y U^k es definida positiva. Sea ahora θ^k el menor valor propio de $U^k, \forall k \in \mathbb{N}$. Entonces $\theta^k = \min_{\|v\|=1} v^t U^k v$.

Afirmación la secuencia $\{\theta^k\}$ es acotada. En efecto, la acotación inferior se tiene del hecho que U^k es definida positiva. Por lo tanto, $\theta^k \geq 0, \forall k \in \mathbb{N}$. Por otra parte, por la proposición 5.4 y por las hipótesis 4.4 y 4.5, las matrices $A^k, (B^k)^{-1}$ y $(\Lambda^k)^{-1}$ son acotadas, y además, por construcción G^k es acotada superiormente. Estos hechos implican que θ^k que sea acotada superiormente. Por lo tanto, existe $N \subseteq \mathbb{N}$ tal que $\{\theta^k\}_{k \in \mathbb{N}}$ converge para algún θ^* . Afirmación $\theta^* > 0$. Para cada $k \in \mathbb{N}$, considere el problema:

$$\begin{cases} \min & v^t (A^k)^t (B^k)^{-1} A^k v - v^t (\Lambda^k)^{-1} G^k v \\ \text{s.a} & g_i^k(x, z) \leq 0, \quad i = 1 \dots, l \\ & \|v\| = 1. \end{cases}$$

Haciendo $k \rightarrow \infty, k \in N$, el problema anterior resulta

$$\min_{\|v\|=1} \left\{ v^t (A^*)^t (B^*)^{-1} A^* v + \sum_{j \in J^*} \frac{[-g_j^*]}{\lambda_j^*} v_j^2 \right\}$$

donde $J^* = \{i \in \{1, \dots, l\} \mid g_i^*(x^*, z^*) = 0\}$, A^* es punto de acumulación de A^k, B^* es un punto de acumulación de $\{B^k\}_{k \in \mathbb{N}}, \lambda^*$ es un punto de acumulación de $\{\lambda^k\}_{k \in \mathbb{N}}$ y $g_j^* = g_j^*(x^*, z^*)$. Ahora, si $\sum_{j \in J^*} \frac{[-g_j^*]}{\lambda_j^*} v_j^2 > 0$, entonces $\theta^* > 0$. Caso contrario $\sum_{j \in J^*} \frac{[-g_j^*]}{\lambda_j^*} v_j^2 = 0$, entonces $v_j = 0, \forall j \in J^*$ y consecuentemente $A^* v = \sum_{i \in I} A_i^* v_i$. Pero por la proposición 5.4, podemos concluir que $\|A^* v\| > \tau_2$ para algún $\tau_2 > 0$. Por lo tanto, en este caso $\theta^* = \min_{\|v\|=1} \{v^t (A^*)^t (B^*)^{-1} A^* v\} > \beta \tau_2^2 > 0$, donde β es el menor valor propio de $(B^*)^{-1}$. Por lo tanto, sea cual fuera el caso, siempre se tiene que $\theta^* > 0$. \square

Proposición 5.5. *El vector $\bar{\lambda}^k = \lambda_{\alpha_l}^k + \rho_{\alpha_l} \lambda_2^k$ es acotado, $\forall k \in \mathbb{N}$.*

Prueba: Como consecuencia del lema 5.4 tenemos que la matriz

$$(U^k)^{-1} = [(J\tilde{g}_l^k)^t(B^k)^{-1}J\tilde{g}_l^k - (\Lambda^k)^{-1}G^k]^{-1}$$

es acotada superiormente. Del algoritmo FDIPA, tomando como el sistema lineal

$$\begin{cases} B^k d_{\alpha_l}^k + J\tilde{g}_l^k \tilde{\lambda}_{\alpha_l}^k &= -\nabla\psi \\ \tilde{\Lambda}_l^k (J\tilde{g}_l^k)^t d_{\alpha_l}^k + G^k \lambda_{\alpha_l}^k &= 0 \end{cases}$$

De la primera ecuación se tiene

$$d_{\alpha_l}^k = -(B^k)^{-1}(\nabla\psi + J\tilde{g}_l^k \tilde{\lambda}_{\alpha_l}^k) \quad (5.12)$$

y de la segunda ecuación

$$(J\tilde{g}_l^k)^t d_{\alpha_l}^k = -(\tilde{\Lambda}_l^k)^{-1} \tilde{G}_l^k \lambda_{\alpha_l}^k \quad (5.13)$$

Reemplazando la ecuación (5.12) en (5.13) obtenemos

$$-(J\tilde{g}_l^k)^t (B^k)^{-1}(\nabla\psi + J\tilde{g}_l^k \tilde{\lambda}_{\alpha_l}^k) = -(\tilde{\Lambda}_l^k)^{-1} \tilde{G}_l^k \lambda_{\alpha_l}^k$$

y además

$$\tilde{\lambda}_{\alpha_l}^k = -[(J\tilde{g}_l^k)^t (B^k)^{-1} J\tilde{g}_l^k - (\tilde{\Lambda}_l^k)^{-1}]^{-1} (J\tilde{g}_l^k)^t (B^k)^{-1} \nabla\psi.$$

Para determinar $\tilde{\lambda}_{\beta_l}^k$ usamos las ecuaciones (5.5) y (5.6) del algoritmo NFDA

$$\tilde{\lambda}_{\beta_l}^k = [(J\tilde{g}_l^k)^t (B^k)^{-1} J\tilde{g}_l^k - (\tilde{\Lambda}_l^k)^{-1} \tilde{G}_l^k]^{-1}.$$

Luego

$$\tilde{\lambda}_{\alpha_l}^k = -[(J\tilde{g}_l^k)^t (B^k)^{-1} J\tilde{g}_l^k - (\tilde{\Lambda}_l^k)^{-1}]^{-1} (J\tilde{g}_l^k)^T (B^k)^{-1} \nabla\psi \quad \text{y} \quad \tilde{\lambda}_{\beta_l}^k = (U^k)^{-1}$$

y como ρ_k es acotado tendremos que $\bar{\lambda}^k$ es acotada. □

Lema 5.5. *Para cualquier $(x, z) \in \text{int}(\text{epi}(f))$ y cualquier dirección d dado por el algoritmo NFDA, existe un $\tau > 0$ tal que $\tilde{g}_i((x, z) + td) \leq 0$, $\forall t \in [0, \tau]$, $\forall i = 1, \dots, l$.*

Prueba: Sea b un vector tal que $b_i = s_i^t x_i - f(x_i)$ para cualquier $i = 0, 1, \dots, l$.

Entonces, $\tilde{g}_l((x, z) + td) = (J\tilde{g}_l(x, z))^t \cdot (x, z) - b$, desde que

$$\begin{aligned} g_i(x, z) &= f(y_i) + s_i^t(x - y_i) - z \\ &= \langle (s_i^t, -1) \cdot (x, z) \rangle - b_i \\ &= (\nabla g_i(x, z))^t (x, z) - b_i \end{aligned}$$

para cualquier $i = 0, 1, \dots, l$. La longitud de paso t es definida por la ecuación (5.7) en el paso 3) del algoritmo NFDA. Desde que las restricciones del problema (AP_l) son lineales, para satisfacer las condición de búsqueda lineal, la siguiente desigualdad debe ser verdadera:

$$\begin{aligned} g_i((x, z) + td) &= (\nabla g_i((x, z) + td))^t \cdot ((x, z) + td) - b_i \\ &= g_i(x, z) + t(\nabla g_i(x, z))^t d \leq 0 \end{aligned} \quad (5.14)$$

para cualquier $i = 0, 1, \dots, l$. Si $(\nabla g_i(x, z))^t d \leq 0$ la desigualdad es satisfecha para cualquier $t > 0$, entonces $g_i((x, z) + td) \leq 0$. Otro caso, se sigue de (iii) en el paso 2) que,

$$(\nabla g_i(x, z))^t d = (\nabla g_i(x, z))^t (d_\alpha + \rho d_\beta)$$

Se sigue de la ecuación (5.4) y de la ecuación (5.6) que

$$(\nabla g_i(x, z))^t d_\alpha = -g_i(x, z) \frac{\lambda_{\alpha i}}{\lambda_i}$$

y

$$(\nabla g_i(x, z))^t d_\beta = -1 - g_i(x, z) \frac{\lambda_{\beta i}}{\lambda_i}.$$

Entonces, de la ecuación (5.14), es equivalente a

$$g_i(x, z) \left(1 - t \frac{\bar{\lambda}_i}{\lambda_i} \right) - \rho t_i \leq 0,$$

donde $\bar{\lambda} = \lambda_\alpha + \rho \lambda_\beta$. Como $\rho t > 0$, la última desigualdad es siempre verdadera cuando $t \leq \lambda_i / \bar{\lambda}_i$.

De la hipótesis 4.4, $\lambda > 0$ es acotada y tenemos que $\bar{\lambda}$ es acotada superiormente. Así,

existe $0 < \tau < t_{max}/\mu$ tal que $\tau < \lambda_i/\bar{\lambda}_i$ para cualquier $i = 0, 1, \dots, l$. Por lo tanto, para cualquier $t \in [0, \tau]$ la condición de búsqueda lineal $g_i((x, z) + td) \leq 0$ es satisfecha para cualquier $i = 0, 1, \dots, l$. \square

Lema 5.6. *Sea λ un valor propio de la matriz $A(A^t B^{-1} A - \Lambda^{-1} G)^{-1} A^t B^{-1}$. Entonces $\lambda \leq 1$.*

Prueba: Sea u un vector propio asociado a λ

$$A(A^t B^{-1} A - \Lambda^{-1} G)^{-1} A^t B^{-1} u = \lambda u$$

Si $A^t B^{-1} u = 0$ entonces $\lambda = 0$. Caso contrario, si $A^t B^{-1} u = v$, tenemos

$$A^t B^{-1} A(A^t B^{-1} A - \Lambda^{-1} G)^{-1} v = \lambda v \quad (5.15)$$

Pero

$$\begin{aligned} A^t B^{-1} A(A^t B^{-1} A - \Lambda^{-1} G)^{-1} &= [(A^t B^{-1} A - \Lambda^{-1} G) + \Lambda^{-1} G](A^t B^{-1} A - \Lambda^{-1} G)^{-1} \\ &= I + \Lambda^{-1} G(A^t B^{-1} A - \Lambda^{-1} G)^{-1} \end{aligned}$$

de esta última expresión y de la ecuación (5.15) se sigue

$$[I + \Lambda^{-1} G(A^t B^{-1} A - \Lambda^{-1} G)^{-1}] v = \lambda v \quad (5.16)$$

Multiplicando a la ecuación (5.16) por $v^t(A^t B^{-1} A - \Lambda^{-1} G)^{-1}$ obtenemos

$$\begin{aligned} v^t(A^t B^{-1} A - \Lambda^{-1} G)^{-1} v + v^t(A^t B^{-1} A - \Lambda^{-1} G)^{-1} \Lambda^{-1} G(A^t B^{-1} A - \Lambda^{-1} G)^{-1} v \\ = \lambda v^t(A^t B^{-1} A - \Lambda^{-1} G)^{-1} v \end{aligned}$$

o

$$v^t(A^t B^{-1} A - \Lambda^{-1} G)^{-1} \Lambda G(A^t B^{-1} A - \Lambda^{-1} G)^{-1} v = (\lambda - 1) v^t(A^t B^{-1} A - \Lambda^{-1} G)^{-1} v \quad (5.17)$$

Desde que $(A^t B^{-1} A - \Lambda^{-1} G)^{-1}$ es definida positiva, concluimos de la ecuación (5.17) que $\lambda \leq 1$. \square

Lema 5.7. Las secuencias $\{d_\alpha^k\}_{k \in \mathbb{N}}$ y $\{d_\beta^k\}_{k \in \mathbb{N}}$ son acotadas.

Prueba: Sabemos que

$$d_\alpha^k = -(B^k)^{-1}(\nabla\psi + J\tilde{g}_l^k \lambda_\alpha^k) \quad (5.18)$$

$$\lambda_\alpha^k = -((J\tilde{g}_l^k)^t(B^k)^{-1}J\tilde{g}_l^k - (\Lambda^k)^{-1}G^k)^{-1}(J\tilde{g}_l^k)^{-1}(J\tilde{g}_l^k)^t(B^k)^{-1}\nabla\psi \quad (5.19)$$

$$d_\beta^k = -(B^k)^{-1}J\tilde{g}_l^k \lambda_\beta^k \quad (5.20)$$

$$\lambda_\beta^k = [(J\tilde{g}_l^k)^t(B^k)^{-1}J\tilde{g}_l^k - (\Lambda^k)^{-1}G^k]^{-1} \quad (5.21)$$

Reemplazando la ecuación (5.19) en (5.18) obtenemos

$$d_\alpha^k = -(B^k)^{-1}[I - J\tilde{g}_l^k((J\tilde{g}_l^k)^t(B^k)^{-1}J\tilde{g}_l^k - (\Lambda^k)^{-1}G^k)^{-1}(J\tilde{g}_l^k)^t(B^k)^{-1}]\nabla\psi$$

Consecuentemente:

$$\|d_\alpha^k\| \leq \|(B^k)^{-1}\| [1 + \|J\tilde{g}_l^k((J\tilde{g}_l^k)^t(B^k)^{-1}J\tilde{g}_l^k - (\Lambda^k)^{-1}G^k)^{-1}(J\tilde{g}_l^k)^t(B^k)^{-1}\|] \|\nabla\psi\|$$

Siendo $(B^k)^{-1}$ acotada y considerando el lema 5.6, concluimos que $\|d_\alpha^k\| < \infty$. Para probar que la secuencia $\{d_\beta^k\}$ sea acotada, reemplazamos la ecuación (5.21) en (5.20) obteniendo

$$d_\beta^k = -(B^k)^{-1}J\tilde{g}_l^k[(J\tilde{g}_l^k)^t(B^k)^{-1}J\tilde{g}_l^k - (\Lambda^k)^{-1}G^k]^{-1}$$

Considerando las hipótesis 4.5, 4.6 y el lema 5.4, concluimos que $\|d_\beta^k\| < \infty$. \square

Proposición 5.6. Sea $(s, -1)^t$ el gradiente de una restricción activa en (x^*, z^*) . Entonces $s \in \partial f(x^*)$.

Prueba: Sea $g(x, z) = f(u) + \langle s, x - u \rangle - z \leq 0$ una restricción activa en (x^*, z^*) donde $s \in \partial f(u)$. Entonces

$$g(x^*, z^*) = f(u) + \langle s, x^* - u \rangle - z^* = 0 \quad (5.22)$$

Como f es convexa podemos escribir

$$\begin{aligned} f(x) &\geq f(u) + \langle s, x - u \rangle = f(u) - f(x^*) + f(x^*) + \langle s, x - u \rangle - \langle s, x - x^* \rangle + \langle s, x - x^* \rangle \\ &= f(x^*) + \langle s, x - x^* \rangle + f(u) - f(x^*) + \langle s, x - u \rangle + \langle s, x^* - x \rangle \\ &= f(x^*) + \langle s, x - x^* \rangle + [f(u) - f(x^*) + \langle s, x^* - u \rangle]. \end{aligned}$$

y por (5.22), $f(x^*) = z^*$. Luego $f(x) \geq f(x^*) + \langle s, x - x^* \rangle \Rightarrow s \in \partial f(x^*)$. \square

Proposición 5.7. *Sea d_α^* un punto de acumulación de la secuencia $\{d_\alpha^k\}_{k \in \mathbb{N}}$. Entonces $d_\alpha^* = 0$.*

Prueba: Del paso 4) del algoritmo NFDA, se tiene

$$(x^{k+1}, z^{k+1}) = (x^k, z^k) + \mu t^k (d_x^k, d_z^k)$$

Así

$$z^{k+1} = z^k + \mu t^k d_z^k \tag{5.23}$$

y la secuencia $\{z^k\}_{k \in \mathbb{N}}$ es decreciente y pertenece al conjunto compacto Ω_a . Denote por $z^* = \lim_{k \rightarrow \infty} z^k$ y $\mathbb{N}' \subset \mathbb{N}$ tal que $\{t^k\}_{k \in \mathbb{N}'} \rightarrow t^*$. Se sigue del lema 5.5. que $t^* > 0$, cuando $k \rightarrow \infty$, $k \in \mathbb{N}'$ de (5.23) tenemos $z^* = z^* + \mu t^* d_z^*$, así $d_z^* = 0$. Por la proposición 5.1. se sigue

$$0 = d_z^* \leq \xi (d_\alpha^k)^t \nabla \psi(x, z) = \xi d_{\alpha z}^* \leq 0, \text{ así } d_{\alpha z}^* = 0,$$

por lo tanto por el lema 5.2. tenemos

$$0 = d_{\alpha z}^* = (d_\alpha^*)^t \nabla \psi(x, z) \leq -(d_\alpha^*)^t B d_\alpha^* \leq 0$$

Así, $d_\alpha^* = 0$, desde que B es una matriz definida positiva. \square

De los resultados previos se tiene que

$$d^k = d_\alpha^k + \rho_k d_\beta^k \rightarrow 0, \quad k \rightarrow \infty, \quad k \in \mathbb{N}'$$

desde que $\rho_k \rightarrow 0$ y $d_\alpha^k \rightarrow 0$. Consecuentemente el proceso puede parar cuando $\{d_\alpha^k\}$

converge hacia cero.

Proposición 5.8. *Para cualquier punto de acumulación (x^*, z^*) de la secuencia $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ generada por el algoritmo NFDA, se tiene $0 \in \partial f(x^*)$, es decir, x^* es una solución del problema (P).*

Prueba: Denote por:

$$\mathbb{Y} = \{y_0^0, y_1^0, \dots, y_{l^0}^0, y_0^1, y_1^1, \dots, y_{l^1}^1, \dots, y_0^k, y_1^k, \dots, y_{l^k}^k, \dots\},$$

la secuencia de todos los puntos obtenidos en las sub-iteraciones por el algoritmo NFDA. Dado que esta secuencia es acotada, existirá una subcolección de ella la cual será convergente. En particular, desde que $x^{k+1} = y_{l^k}^k$, existirá un $\bar{\mathbb{Y}} \subset \mathbb{Y}$ tal que $\{y^k\}_{k \in \bar{\mathbb{Y}}}$ converja a x^* .

Definiendo $\bar{\mathbb{Y}}^k = \{y_0^k, y_1^k, \dots, y_{l^k}^k\} \cap \bar{\mathbb{Y}}$. Se tiene que, de la ecuación (5.3) del paso 2. del algoritmo NFDA, $B^k d_{\alpha l} + J\tilde{g}_l(x^k, z^k)\tilde{z}_{\alpha l}^k = -\nabla\psi(x^k, z^k)$.

$$\lim_{k \rightarrow \infty} \sum_{i=0}^{l^k} \lambda_{\alpha i}^k s_i^k = 0 \quad y \quad \lim_{k \rightarrow \infty} \sum_{i=0}^{l^k} \lambda_{\alpha i}^k = 1. \quad (5.24)$$

Definimos ahora el conjunto de índices $\mathbb{I}^k = \{i \mid y_i^k \in \bar{\mathbb{Y}}^k\}$. Consecuentemente $\lambda_{\alpha i}^k > 0$ para $i \in \mathbb{I}^k$ y k suficientemente grande. Entonces,

$$\lim_{k \rightarrow \infty} \sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k s_i^k = 0 \quad y \quad \lim_{k \rightarrow \infty} \sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k = 1 \quad (5.25)$$

Sea el punto auxiliar y_i^k y el subgradiente $s_i^k \in \partial f(y_i^k)$ de manera que $i \in \mathbb{I}^k$. Luego, podemos escribir

$$f(x) \geq f(y_i^k) + (s_i^k)^t \cdot (x - y_i^k) = f(x^k) + (s_i^k)^t \cdot (x - x^k) - \epsilon_k^i$$

donde

$$\epsilon_k^i = f(x^k) - f(y_i^k) - (s_i^k)^t \cdot (x^k - y_i^k)$$

Como $x^k \rightarrow x^*$ y $y_i^k \rightarrow x^*$ para $i \in \mathbb{I}^k$, tenemos que $\lim_{k \rightarrow \infty} \epsilon_k^i = 0$. Por tanto $s_i^k \in \partial_{\epsilon_k^i} f(x^k)$. Luego

$$\left(\sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k \right) f(x) \geq \left(\sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k \right) f(x^k) + \left(\sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k s_i^k \right) \cdot (x - x^k) - \sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k \epsilon_k^i,$$

$$f(x) \geq f(x^k) + \left(\sum_{i \in \mathbb{I}^k} \frac{\lambda_{\alpha i}^k}{\sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k} s_i^k \right) (x - x^k) - \epsilon_k \quad (5.26)$$

donde $\epsilon_k = \frac{\sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k \epsilon_k^i}{\sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k}$ y como $\lim_{k \rightarrow \infty} \epsilon_k^i = 0$ y $\lim_{k \rightarrow \infty} \sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k = 1$, tenemos que $\lim_{k \rightarrow \infty} \epsilon_k = 0$. Considerando $s^k = \sum_{i \in \mathbb{I}^k} \frac{\lambda_{\alpha i}^k}{\sum_{i \in \mathbb{I}^k} \lambda_{\alpha i}^k} s_i^k$, se tiene que, de la ecuación (5.26) que $s^k \in \partial_{\epsilon_k} f(x^k)$ y $0 \in \partial f(x^*)$. Por lo tanto, esto completa la prueba de la convergencia. \square

5.4. Algunas ideas para solucionar otros tipos de problemas con el método NFDA

Consideremos el siguiente problema de optimización convexa

$$(P_1) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s.a} & g(x) \leq 0 \end{cases}$$

donde la función objetivo y la restricción de desigualdad $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$ son convexas y no necesariamente diferenciables.

Reemplazando el problema a su equivalente diferenciable lineal, se tiene

$$(EP_1) \quad \begin{cases} \text{minimizar} & \psi(x, t) = t \\ \text{s.a} & f(x) \leq t \\ & g(x) \leq 0 \end{cases}$$

Luego, se aplica a cada restricción con la misma técnica usada en el método NFDA, encontrando un problema diferenciable con restricción diferenciable y procedemos como en el método NFDA.

Si podemos generamos una secuencia $\{x^k\}$ donde sus puntos de acumulación converjan a la solución del problema (P) , entonces podríamos generalizar a un problema cuyas restricción de desigualdad sea la función $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

De igual manera si consideramos el siguiente problema de optimización no diferenciable con restricciones de igualdad

$$(P_2) \quad \begin{cases} \text{minimizar} & f(x) \\ \text{s.a} & h(x) = 0 \end{cases}$$

donde la función objetivo y la restricción de igualdad $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ son funciones continuas.

Ahora definiendo su función lagrangiana $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+ \rightarrow \mathbb{R}$ asociada al problema (P_2) por

$$L(x, u, c) = f(x) + c\sigma(h(x)) + \langle Au, h(x) \rangle,$$

donde A es una matriz de orden $n \times m$ simétrica y $\sigma : \mathbb{R}^m \rightarrow \mathbb{R}$ es una función continua tal que $\sigma(z) > 0, \forall z \in \mathbb{R}^m - \{0\}$ y $\sigma(0) = 0$, para ver más detalles de la función lagrangiana ver por ejemplo [10] y [11]. Luego la función dual $H : \mathbb{R}^m \times \mathbb{R}_+ \rightarrow \mathbb{R}$ es definida por

$$H(u, c) = \inf_{x \in \mathbb{R}^n} [f(x) + c\sigma(h(x)) + \langle Au, h(x) \rangle]. (\text{cóncavo})$$

Entonces el problema dual es dado por

$$(D) \quad \begin{cases} \text{minimizar} & -H(u, c) \\ \text{s.a} & (u, c) \in \mathbb{R}^m \times \mathbb{R}_+. \end{cases}$$

Luego este problema lo llevamos a su equivalente diferenciable, es decir,

$$(EP_2) \quad \begin{cases} \text{minimizar} & \psi(u, c) = t \\ \text{s.a} & H(u, c) \leq t, \end{cases}$$

y aplicamos la misma técnica usada en el método NFDA para resolver el problema (P_2) .

Si generamos una secuencia $\{x^k\}$ podríamos probar que cada punto de acumulación es una solución del problema (P_2) .

Si en el problema anterior tuviéramos ahora una restricción de desigualdad $g : \mathbb{R}^n \rightarrow \mathbb{R}^q$, esta también podría ser resuelta sustituyendo la restricción por la función $\bar{g}_i(x) = \max\{g_i(x), 0\}$ y así aplicamos el mismo criterio realizado en el problema (P_2) . Otra forma sería usando una función lagrangiana donde incluiríamos la restricción de desigualdad.

Capítulo 6

Resultados numéricos

El algoritmo fue implementado en el software libre Octave 6.2.0 sobre una laptop ASUS con un microprocesador Intel(R) Core(TM) i5-6200U, de 2.3GHz y 2.40 GHZ con memoria 8.00 GB de RAM. Para la verificación de la velocidad de convergencia y en dar una buena aproximación a la solución óptima con el método NFDA, fueron tomados algunos problemas test de ([24], [22], [27], [34],[37], [29]). Estos problemas test son presentados a continuación, los cuales varían de 2 a 100 variables y con 1 a 48 restricciones. Fueron extraídos del texto [28] otros métodos convexos no diferenciables, para realizar la comparación de la velocidad y exactitud con el método NFDA.

6.1. Problemas Test

Crescent [19]

Dimensión 2

Función objetivo $f(x) = \max\{x_1^2 + (x_2 - 1)^2 + x_2 - 1, -x_1^2 - (x_2 - 1)^2 + x_2 + 1\}$

Punto óptimo $x^* = [0 \ 0]$

Valor óptimo $f(x^*) = 0$

Punto inicial $x = [-1,5 \ 2]$

CB2 [12]

Dimensión 2

Función objetivo $f(x) = \max\{x_1^2 + x_2^4, (2 - x_1)^2 + (2 - x_2)^2, 2e^{-x_1+x_2}\}$ Punto óptimo $x^* = [1,139286 \ 0,899365]$ Valor óptimo $f(x^*) = 1,952225$ Punto inicial $x = [1 \ -0,1]$ **CB3 [12]**

Dimensión 2

Función objetivo $f(x) = \max\{x_1^4 + x_2^2, (2 - x_1)^2 + (2 - x_2)^2, 2e^{-x_1+x_2}\}$ Punto óptimo $x^* = [1 \ 1]$ Valor óptimo $f(x^*) = 2$ Punto inicial $x = [2 \ 2]$ **DEM [29]**

Dimensión 2

Función objetivo $f(x) = \max\{5x_1 + x_2, -5x_1 + x_2, x_1^2 + x_2^2 + 4x_2\}$ Punto óptimo $x^* = [0 \ -3]$ Valor óptimo $f(x^*) = -3$ Punto inicial $x = [1 \ 1]$

QL [29]

Dimensión	2
Función objetivo	$f(x) = \max_{1 \leq i \leq 3} f_i(x)$
Donde	$f_1(x) = x_1^2 + x_2^2$ $f_2(x) = x_1^2 + x_2^2 + 10(-4x_1 - x_2 + 4)$ $f_3(x) = x_1^2 + x_2^2 + 10(-x_1 - 2x_2 + 6)$
Punto óptimo	$x^* = [1,2 \ 2,4]$
Valor óptimo	$f(x^*) = 7,2$
Punto inicial	$x = [-1 \ 5]$

LQ [29]

Dimensión	2
Función objetivo	$f(x) = \max\{-x_1 - x_2, -x_1 - x_2 + (x_1^2 + x_2^2 - 1)\}$
Punto óptimo	$x^* = \left[\frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}} \right]$
Valor óptimo	$f(x^*) = -\sqrt{2}$
Punto inicial	$x = [-0,5 \ -0,5]$

Mifflin1 [29]

Dimensión	2
Función objetivo	$f(x) = -x_1 + 20 \max\{x_1^2 + x_2^2 - 1, 0\}$
Punto óptimo	$x^* = [1 \ 0]$
Valor óptimo	$f(x^*) = -1$
Punto inicial	$x = [0,8 \ 0,6]$

Mifflin2 [29]

Dimensión	2
Función objetivo	$f(x) = -x_1 + 2(x_1^2 + x_2^2 - 1) + 1,75 x_1^2 + x_2^2 - 1 $
Punto óptimo	$x^* = [1 \ 0]$
Valor óptimo	$f(x^*) = -1$
Punto inicial	$x = [-1 \ -1]$

EVD52 [27]

Dimensión	3
Función objetivo	$f(x) = \max_{1 \leq i \leq 6} f_i(x)$
	$f_1(x) = x_1^2 + x_2^2 + x_3^3 - 1$
	$f_2(x) = x_1^2 + x_2^2 + (x_3 - 2)^2$
	$f_3(x) = x_1 + x_2 + x_3 - 1$
	$f_4(x) = x_1 + x_2 - x_3 - 1$
	$f_5(x) = 2x_1^3 + 6x_2^2 + 2(5x_3 - x_1 + 1)^2$
	$f_6(x) = x_1^2 - 9x_3$
Punto óptimo	$x^* = [0,3283 \ 0,0000 \ 0,1313]$
Valor óptimo	$f(x^*) = 3,5997193$
Punto inicial	$x = [1 \ 1 \ 1]$

Rosen Suzuki [34]

Dimensión	4
Función objetivo	$f(x) = \max\{f_1(x), f_1(x) + 10f_2(x)f_1(x) + 10f_3(x), f_1(x) + 10f_4(x)\}$
Donde	$f_1(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4$ $f_2(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8$ $f_3(x) = x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10$ $f_4(x) = x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5$
Punto óptimo	$x^* = [0, 1 \ 1 \ 2 \ -1]$
Valor óptimo	$f(x^*) = -44$
Punto inicial	$x = [0 \ 0 \ 0 \ 0]$

Davidon2 [27]

Dimensión	4
Función objetivo	$f(x) = \max_{1 \leq i \leq 20} \{f_i(x)\}$ $f_i(x) = (x_1 + x_2 t_i - e^{t_i})^2 + (x_3 + x_4 \text{sen}(t_i) - \text{cos}(t_i))^2$ $t_i = 0,2i, \text{ para } i = 1, \dots, 20$
Punto óptimo	$x^* = [-12,2437 \ 14,0218 \ -0,4515 \ -0,0105]$
Valor óptimo	$f(x^*) = 115,70644$
Punto inicial	$x = [25 \ 5 \ -5 \ -1]$

Shor [29]

Dimensión 5

Función objetivo $f(x) = \max_{1 \leq i \leq 10} \{b_i \sum_{j=1}^5 (x_j - a_{ij})^2\}$ Punto óptimo $x^* = [1 \ 1 \ 1 \ 1]$ Valor óptimo $f(x^*) = 22,60016$ Punto inicial $x = [0 \ 0 \ 0 \ 0 \ 1]$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 1 & 1 & 3 \\ 1 & 2 & 1 & 1 & 2 \\ 1 & 4 & 1 & 2 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 0 & 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 1 \\ 0 & 0 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 \end{pmatrix}; \quad b = \begin{pmatrix} 1 \\ 2 \\ 10 \\ 2 \\ 4 \\ 3 \\ 1,7 \\ 2,5 \\ 6 \\ 3,5 \end{pmatrix}$$

Maxquad [24]

Dimensión 10

Función objetivo $f(x) = \max_{1 \leq i \leq 5} \{x^t A_i x - b_i^t x\}$ Valor óptimo $f(x^*) = -0,8414084$ Punto inicial 1 $x = [1 \ \dots \ 1]$ Punto inicial 2 $x = [0 \ \dots \ 0]$

Donde

$$A_{ijk} = A_{ikj} = e^{\frac{i}{k}} \cos(jk) \text{sen}(i), \quad j < k$$

$$A_{jjj} = \frac{j}{10} |\text{sen}(i)| + \sum_{k \neq j} |A_{ijk}| \quad y$$

$$b_{ij} = e^{\frac{i}{j}} \text{sen}(ij)$$

MaxQ [37]

Dimensión 20

Función objetivo $f(x) = \max_{1 \leq i \leq 20} \{x_i^2\}$ Valor óptimo $f(x^*) = 0$ Punto inicial 1 $x_i = i, i = 1, \dots, 10$
 $x_i = -i, i = 1, \dots, 20$ **MaxL [29]**

Dimensión 20

Función objetivo $f(x) = \max_{1 \leq i \leq 20} |x_i|$ Valor óptimo $f(x^*) = 0$ Punto inicial $x_i = i, i = 1, \dots, 10$
 $x_i = -i, i = 11, \dots, 20.$

Wong 3 [32]

Dimensión 20

Función objetivo $f(x) = \max_{1 \leq i \leq 18} \{f_i\}$

$$f_1(x) = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 \\ + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 \\ + (x_{10} - 7)^2 + (x_{11} - 9)^2 + 10(x_{12} - 1)^2 + 5(x_{13} - 7)^2 \\ + 4(x_{14} - 14)^2 + 27(x_{15} - 1)^2 + x_{16}^4 + (x_{17} - 2)^2 \\ + 13(x_{18} - 2)^2 + (x_{19} - 3)^2 + x_{20}^2 + 95,$$

$$f_2(x) = f_1(x) + 10(3(x_1 - 2)^2 + 4(x_2 - 3)^2 + 2x_3^2 - 7x_4 - 120),$$

$$f_3(x) = f_1(x) + 10(5x_1^2 + 8x_2 + (x_3 - 6)^2 - 2x_4 - 40)$$

$$f_4(x) = f_1(x) + 10(0,5(x_1 - 8)^2 + 2(x_2 - 4)^2 + 3x_5^2 - x_6 - 30)$$

$$f_5(x) = f_1(x) + 10(x_1^2 + 2(x_2 - 2)^2 - 2x_1x_2 + 14x_5 - 6x_6),$$

$$f_6(x) = f_1(x) + 10(4x_1 + 5x_2 - 3x_7 + 9x_8 - 105),$$

$$f_7(x) = f_1(x) + 10(10x_1 - 8x_2 - 17x_7 + 2x_8),$$

$$f_8(x) = f_1(x) + 10(-3x_1 + 6x_2 + 12(x_9 - 8)^2 - 7x_{10}),$$

$$f_9(x) = f_1(x) + 10(-8x_1 + 2x_2 + 5x_9 - 2x_{10} - 12),$$

$$f_{10}(x) = f_1(x) + 10(x_1 + x_2 + 4x_{11} - 21x_{12}),$$

$$f_{11}(x) = f_1(x) + 10(x_{21} + 15x_{11} - 8x_{12} - 28),$$

$$f_{12}(x) = f_1(x) + 10(4x_1 + 9x_2 + 5x_{13}^2 - 9x_{14} - 87),$$

$$f_{13}(x) = f_1(x) + 10(3x_1 + 4x_2 + 3(x_{13} - 6)^2 - 14x_{14} - 10),$$

$$f_{14}(x) = f_1(x) + 10(14x_1^2 + 35x_{15} - 79x_{16} - 92),$$

$$f_{15}(x) = f_1(x) + 10(15x_2^2 + 11x_{15} - 61x_{16} - 54),$$

$$f_{16}(x) = f_1(x) + 10(5x_1^2 + 2x_2 + 9x_{17}^4 - x_{18} - 68),$$

$$f_{17}(x) = f_1(x) + 10(x_1^2 - x_2 + 19x_{19} - 20x_{20} + 19),$$

$$f_{18}(x) = f_1(x) + 10(x_1^2 + 5x_2^2 + x_1^29 - 30x_{20}),$$

Valor óptimo $f(x^*) = 133,72828$ Punto inicial $x_0 = (2, 3, 5, 5, 1, 2, 7, 3, 6, 10, 2, 2, 6, 15, 1, 2, 1, 2, 1, 3)^T$

TR48 [22]

Dimensión 48

$$\text{Función objetivo } f(x) = \sum_{j=1}^{48} d_j \max_{1 \leq i \leq 48} \{x_i - a_{ij}\} - \sum_{i=1}^{48} s_i x_i$$

$$\text{Valor óptimo } f(x^*) = -638565$$

Punto inicial $x_i = 0, i = 1, \dots, 48$

Donde el vector $s =$

22	53	64	15	66	37	16	23	67	18	52	69
17	29	50	13	95	34	59	36	22	94	28	34
36	38	55	77	45	34	32	58	30	88	74	59
93	54	89	30	79	46	35	41	99	52	76	93

vector $d =$

61	67	24	84	13	86	89	46	48	50	74	75
88	40	29	45	32	21	61	21	51	14	89	79
38	20	97	19	10	73	59	92	52	66	89	65
63	47	7	61	87	19	36	43	9	12	8	67

y la matriz simétrica $A = (a_{ij})$;

273	1272	744	1136	1972	1580	1878	1539	1457	429	1129	1251	1421
588	334	837	1364	229	961	734	1169	1488	720	1280	816	664
1178	939	1698	983	1119	1029	1815	721	1753	330	1499	1107	1576
942	484	617	896	1184	1030	1718	604	999	809	866	1722	1338
1640	1266	1185	440	894	992	1173	334	358	626	1124	358	847
533	915	1219	481	1009	543	937	915	667	1441	812	848	776
1560	526	1494	598	1244	1304	1306	685	668	444	1157	1359	1176
1475	335	1519	140	937	697	951	267	227	1229	587	369	554
721	1212	739	596	1291	1114	701	426	285	676	155	456	1936
319	337	604	907	214	424	748	817	666	1592	521	2172	356
467	1583	882	2139	2182	1961	781	678	1425	1861	1473	1713	1761
1617	370	1073	1304	1369	1092	453	798	1283	973	565	1315	1204
1796	846	1447	1143	959	1275	1213	2085	742	1309	1479	1760	703
1727	872	1479	686	1698	1057	387	1252	904	668	443	1600	930
1052	776	1049	402	361	1119	578	406	618	581	1095	670	641
1152	1060	567	433	374	579	235	325	1802	331	217	665	862
182	312	864	732	783	1456	608	2066	491	400	1466	744	2013
2082	1865	875	552	400	182	820	721	1735	851	740	551	1551
1769	1159	613	2072	1300	1605	807	1017	1251	818	1259	2596	826
1137	1255	1123	943	1359	188	1282	271	2300	483	2540	609	1038
2099	1766	2699	2493	2266	264	1398	304	699	538	1335	454	393
173	1198	1370	760	216	1692	919	1286	435	879	861	548	913
2198	483	803	1181	731	627	1086	292	883	279	1906	178	2156
490	662	1699	1430	2300	2117	1888	138	1023	884	755	1612	749
690	476	1501	1654	1049	516	1995	1149	1580	739	1079	1161	815
1214	2485	780	1100	1347	985	916	1361	260	1171	328	2202	445
2385	665	966	1969	1729	2568	2333	2108	177	1327	177	1486	757
506	609	981	1474	967	681	1552	1317	936	594	197	928	316
723	2203	500	604	482	1104	455	630	641	1058	562	1857	528
2425	220	704	1845	1122	2405	2428	2204	738	945	1362	587	335

435	930	1358	819	504	1496	1153	927	428	341	803	180	649
2119	343	521	652	939	340	649	533	918	451	1783	362	2290
130	568	1727	1105	2301	2285	2059	595	853	891	1082	1199	726
96	583	1125	653	563	947	986	1493	560	1183	813	882	1033
902	1763	642	1032	1131	1604	463	1556	663	1298	947	1461	795
371	882	967	973	768	1472	588	252	308	803	920	309	238
1252	569	940	165	863	414	454	552	1745	269	482	1188	355
397	833	713	432	666	1453	410	1758	642	262	1260	1051	1858
1737	1508	592	598	222	814	1094	510	235	1335	820	892	100
626	541	219	524	1897	90	410	952	605	238	706	570	622
503	1581	257	1985	396	309	1453	1039	2043	1972	1744	514	661
1025	1227	617	90	1525	835	1114	263	770	700	400	740	2049
311	630	1087	630	459	924	405	739	360	1749	115	2055	428
492	1568	1256	2166	2026	1796	303	853	663	632	999	572	972
225	763	908	451	767	293	1240	726	420	1111	862	617	443
1374	586	1299	887	1070	1633	1057	547	999	252	1483	1681	1489
1326	236	610	1156	557	642	879	1000	1467	558	1178	780	831
1038	879	1726	700	1023	1082	1631	488	1579	586	1320	982	1463
796	371	802	949	1021	826	1508	550	546	983	397	821	411
1023	180	651	478	1438	476	485	1333	235	525	627	1022	123
973	1155	715	1475	902	273	953	882	1550	1457	1240	898	396
1479	745	1105	240	831	645	442	723	1983	316	623	1152	543
470	939	482	669	443	1690	205	1969	510	455	1492	1238	2091
1938	1709	354	813	1163	676	1264	1473	839	1326	847	801	1254
976	1643	1157	1169	983	1905	878	1836	346	1590	1286	1621	1034
689	503	995	1376	1239	1828	674	1183	725	1399	549	1004	869
1427	818	882	1716	214	902	1222	1210	390	1184	1225	949	239
1210	660	863	1207	1446	1197	969	1042	741	865	821	644	790
388	1374	803	484	968	1056	665	318	1420	794	1341	1017	1137
1836	1056	679	1200	189	1645	1891	1704	1403	442	699	453	290

483	1809	107	384	1024	511	251	712	646	525	585	1499	330
1885	495	231	1356	999	1949	1872	1644	567	591	950	410	690
2147	594	590	326	1191	499	504	838	1098	758	1794	703	2439
414	751	1837	1011	2374	2455	2237	928	921	624	325	1356	480
369	1241	413	473	680	1097	166	1038	1049	781	1497	905	238
925	702	1506	1506	1287	998	216	479	1941	188	350	736	792
161	547	632	745	552	1607	375	2115	296	392	1547	959	2121
2114	1890	641	676	1480	435	129	949	708	325	355	1081	492
1007	1137	779	1759	774	291	1148	516	1688	1785	1573	1038	231
1829	1603	2339	1524	1780	1673	2421	1315	2394	357	2136	825	2237
1589	579	1204	347	959	940	2336	1266	320	919	605	154	623
652	580	582	1508	344	1950	429	242	1402	949	1986	1943	1717
603	582	872	699	197	358	957	529	881	1263	660	1849	645
240	1250	631	1802	1867	1650	923	341	1511	815	669	1092	1397
1019	1982	1010	2708	695	1061	2089	1148	2594	2734	2520	1212	1176
697	1051	1018	290	985	1280	743	1427	998	466	987	1110	1584
1395	1166	861	626	469	761	607	685	1446	472	1969	457	254
1393	823	1963	1975	1752	739	515	1171	847	1069	1316	919	2063
776	598	1434	507	1926	2101	1898	1187	548	1144	83	2145	317
2445	426	875	1972	1584	2571	2408	2179	194	1231	1094	1036	836
1371	1008	354	833	828	1429	1369	1146	1021	352	2083	259	2412
345	811	1925	1507	2523	2380	2151	220	1163	1828	1005	1903	1272
504	649	653	1114	1019	2044	932	2165	330	559	1668	1291	2264
2138	1908	268	917	2377	1723	636	1720	534	145	290	2281	1531
667	1829	1235	2410	2367	2139	519	972	1162	792	1744	1724	1500
796	381	1087	600	701	550	1835	917	1490	1767	1614	1553	486
678	727	2435	1461	229	2238	1560	2010	1353	1157			

Goffin [29]

Dimensión 50

Función objetivo $f(x) = 50 \max_{1 \leq i \leq 50} \{x_i\} - \sum_{i=1}^{50} x_i$ Valor óptimo $f(x^*) = 0$ Punto inicial $x_i = i - 25,5, i = 1, \dots, 50$ **Badguy**

Dimensión 100

Función objetivo $f(x) = \max\{\|x\| + 2\lambda - 1, 0\}, \lambda \in \left(0, \frac{1}{2}\right)$ Valor óptimo $f(x^*) = 0$ Punto óptimo $x^* \in B(0, 1 - 2\lambda)$ Punto inicial $x_i = 1, i = 1, \dots, 100$ Parámetro utilizado $\lambda = \frac{1}{8}$

6.2. Resultados obtenidos

Para los resultados obtenidos de la implementación del algoritmo NFDA, fueron tomados los siguientes parámetros y condiciones: $B = \frac{1}{2}I$, $\mu = 0,75$, $\varphi = 0,2$, $\xi = 0,6$.

Problema	Resultados				
	it	cti	itt	nf	f^*
Crescent	16	4	20	21	0.0000
CB2	18	14	32	33	1.9522
CB3	10	10	20	27	2.0000
DEM	15	5	20	28	$0.2486 \cdot 10^{-5}$
QL	19	8	27	24	7.2000
LQ	15	8	23	26	-1.4142
Mifflin1	11	26	37	29	-1.0000
Mifflin2	10	29	39	64	-1.000
EVD52	19	35	54	72	3.5997
Rosen Suzuki	20	39	59	60	-44.0000
Davidon2	34	25	59	65	115.7065
Shor	21	35	56	187	2.6002
Maxquad1	29	154	183	243	-0.8414
Maxquad2	29	205	234	245	-0.8414
MaxQ	57	187	244	59	$0.0000 \cdot 10^{-4}$
MaxL	28	30	58	267	$0.8341 \cdot 10^{-7}$
Wong 3	35	228	206	228	$1.3375610 \cdot 10^2$
TR48	51	160	215	69	-638565
Goffin	22	46	58	19	$0.2456 \cdot 10^{-5}$
Badguy	17	1	18	11	0.0000

Tabla 6.1: it: número de iteraciones factibles, cti : total de pasos nulos, itt : número total de iteraciones, nf: número de veces que la función objetivo y subgradiente fueron calculadas y f^* : valor óptimo encontrado.

Problema	NFDA			M1FC1			BT			PB		
	it	nf	f^*	it	nf	f^*	it	nf	f^*	it	nf	f^*
Crescent	16	21	0.0000	31	93	$2.2 \cdot 10^{-6}$	24	27	$9.4 \cdot 10^{-7}$	32	33	$5.4 \cdot 10^{-7}$
CB2	18	33	1.9522	11	31	1.9522	13	16	1.95222	15	16	1.9522
CB3	10	27	2.0000	14	44	2.001415	13	21	2.0000	15	16	2.0000
DEM	15	21	-2.9997	10	33	-3.000	9	13	-3.0000	7	8	-3.0000
QL	19	28	7.2000	12	30	7.2001	12	17	7.2000	17	18	7.2000
LQ	15	24	-1.4142	16	52	-1.1414	10	11	-1.1414	14	15	-1.1414
Mifflin1	12	26	-1.0000	143	281	-0.9999	49	74	-1.0000	22	23	-1.0000
Mifflin2	13	29	-1.000	30	71	-0.9999	6	13	-1.0000	16	17	-1.0000
Rosen Suzuki	20	72	-44.0000	22	61	-43.9999	22	32	-43.9999	40	41	-43.9999
Shor	21	65	2.6002	21	69	22.6001	29	30	22.6001	26	27	22.6001
Maxquad1	29	187	-0.8414	26	69	-0.8413	45	56	-0.8414	41	42	-0.8414
Maxquad2	29	243	-0.8414	20	54	-0.8413	45	49	-0.8414	33	34	-0.8414
MaxQ	57	245	$0.0000 \cdot 10^{-4}$	144	204	0.0000	125	128	0.0000	158	159	0.0000
MaxL	28	59	$0.8341 \cdot 10^{-7}$	138	213	0.0000	74	84	0.0000	34	35	0.0000
TR48	51	206	-638564.94	163	284	-633625.5	165	179	-638565.0	152	153	-638565.0
Goffin	22	69	$0.2490 \cdot 10^{-5}$	72	194	46.2000	50	53	0.0000	51	52	0.0000

Tabla 6.2: Método NFDA Vs M1FC1 (ϵ -Máximo Descenso), BT (Bundle con Regiones de confianza) y PB (Proximal Bundle). it: número de iteraciones factibles. nf: número de iteraciones que fueron calculados los subgradientes y f . f^* : valor óptimo encontrado.

Conclusiones

En este trabajo se presentó un algoritmo de puntos interiores NFDA, que resuelve un problema de optimización convexa no necesariamente diferenciable sin restricciones. El método NFDA basado en los métodos de Planos Cortantes, el FDIPA y método de Bundle (paso factible y no factible), puede ser aplicado para resolver problemas con restricciones de desigualdad, así como problemas con restricciones mixtas (igualdad y desigualdad) cuya función objetivo es continua.

Se presentó algunos resultados numéricos del método NFDA, cuyos problemas son los llamados problemas test, que sirven para medir la eficacia de los algoritmos. Se solucionó problemas desde dos variables hasta 100 variables que dieron buenos resultados en comparación del valor óptimo real.

Para medir la eficiencia del método NFDA, se comparó sus resultados con los métodos M1FC1 (ϵ -Máximo Descenso), BT (Bundle con Regiones de Confianza) y PB (Bundle Proximal), que fueron tomados en [28].

Se puede mejorar el método NFDA usando la técnica del método de Bundle visto en el capítulo 4 de la tesis, los cuales realizan un control de los planos cortantes, así la manera de escoger la matriz B .

Debido a que el algoritmo NFDA requiere del método FDIPA, en la obtención de una dirección de descenso, podemos decir que la velocidad de convergencia es al menos superlineal.

Bibliografia

- [1] A. IZMAILOV, M. SOLODOV, (2005) *Condições de otimalidade, Elementos de Análisis Convexa e de Dualidade* , Otimização- Vol 1 . Rio de Janeiro. IMPA.
- [2] A. IZMAILOV, M. SOLODOV, (2005) *Otimização, Métodos Computacionais* , Otimização- Vol. 2 . Rio de Janeiro. IMPA.
- [3] BERTSEKAS D., (1999). *Nonlinear Programming*. Second Edition, ISBN 1-886529-00-0, 791 pages.
- [4] BAZARAA, M., SHERALI, H., SHETTY, C., (2006) *Nonlinear Programming, Theory and Algorithms*, 3 ed, John Wiley and Sons, Inc, Hoboken, New Jersey.
- [5] BELTRAN, C., HEREDIA J., (2004). *An Effective Line Search for the Subgradient Method*, Polytechnical University of Catalonia, Barcelona, Spain. November 10.
- [6] BENDSØE, M., (2005) *Optimization of structural topology, shape, and material*. Journal of Industrial Management and Optimization 1, 33-38.
- [7] BENTAL, A., NEMIROVSKI, A., (1997) *Robust truss topology design via semidefinite programming*. SIAM Journal on Optimization, v. 7, n. 4, pp. 991-1016, ISSN: 1052-6234.
- [8] BURACHIK R., YALÇM K., (2009). *A deflected Subgradient Method Using a General Augmented Lagrangian Duality With Implications on Penalty Methods*, oct 24.
- [9] BURACHIK R., WILHELM P. FREIRE, C. YALÇM K., (2012). *Interior Epigraph Directions Method for Nonsmooth and Nonconvex Optimization via Generalized Augmented Lagrangian Duality*, May 26.

- [10] BURACHIK, R., KAYA, C., (1995) *O the absence of duality gap for Lagrangeano-type functions*, Berlin, Springer-Verlag. ISBN: 3-540-59057-9.
- [11] BURACHIK, R., KAYA, C., (2012). *An augmented penalty function method with penalty parameter updates for nonconvex optimization*. *Nonlinear Analysis: Theory Meth.Appl.* 75, 1158?1167.
- [12] CHARALAMBOUS, C., CONN, A., (1978). *An efficient method to solve the min-max problem directly*. *SIAM Journal on Numerical Analysis*, 15(1), 162?187.
- [13] FREUND R., (2004). *Subgradient Optimization, Generalized Programming*. Massachusetts Institute of Technology.
- [14] GOLUB, G., VAN LOAN, C., (1982). *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Third ed. Baltimore, MD, Johns Hopkins University Press. ISBN: 0-8018-5413-X; 0-8018-5414-8.
- [15] HERSKOVITS, J., (1982) *Two-Stage Feasible Directions Algorithm for Non-linear Constrained Optimization*, Research Report No. 103, INRIA, BP 105, 78153 Le Chesnay CEDEX, France.
- [16] HERSKOVITS, J., (1998) *A Feasible Directions Interior Point Technique for Non-linear Optimization*. *JOTA, Journal of Optimization Theory and Applications*, vol. 99, pp. 121-146.
- [17] JOSÉ M., SANDRA A., (1995). *Métodos Computacionais de Otimização*. Departamento de Matemática Aplicada.IMECC-UNICAMP.
- [18] KAPPEL, F., KUNTSEVICH A., (2000) *An implementation of Shor's r-algorithm*, *Comput. Optim. Appl.* 15, 193?205.
- [19] KIWIEL, K., (1990). *Proximity control in bundle methods for convex nondifferentiable minimization*. *Mathematical Programming*, 46, 105?122.
- [20] KLANŠEK, U., PŠUNDER M., (2010). *Solving the Nonlinear Transportation Problem by Global Optimization*. Vilniaus Gedimino Technikos Universitetas, ISSN: 1648-4142 (Print) 1648-3480 (Online) Journal homepage.

- [21] LAGES L., (2010). *Curso de Análisi vol. 2*. 11.ed. Rio de Janeiro, IMPA.
- [22] LEMARÉCHAL, C., MIFFLIN, R., (1978). *Nonsmooth optimization*. **IIASA** proceedings series. Oxford: Pergamon Press.
- [23] LEMARÉCHAL C., (1989). *Nondiferentiable Optimization*, in: handbooks in ORandMS, eds. Menhauser, G.L., et al., Elsevier Publishers, North-Holland, pp. 529-572.
- [24] LEMARÉCHAL, C., SAGASTIZÁBAL, C., (1994). *An approach to variable metric bundle methods*. In J. Henry, J. P. Yvon (Eds.) *Lecture Notes in Control and Information Sciences* (Vol. 197, pp. 144?162). New York: Springer.
- [25] LIMA A.L.D.S, *Uma Estratégia de Descomposição por Relaxação Lagrangeana para Otimização da Programação Diária da Operação de Sistemas Hidrotérmicos com Modelagem Detahlada da Rede Elétrica*, Aplicação ao Sistema brasileiro, Tese(Doutorado), COPPE/UFRJ, Rio de Janeiro, 2007.
- [26] LUENBERGER, D., (1984). *Linear and Nonlinear Programming*, 2 ed., Addison-Wesley Publishing Company.
- [27] LUKŠAN, L., VLČEK, J. *Test problems for nonsmooth unconstrained and linearly constrained optimization*. Technical Report 798, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague.
- [28] MÄKELÄ, M.M., NEITTAANMÄKI, P., (1992). *Nonsmooth optimization: Analysis and algorithms with applications to optimal control*. Department of Mathematics, University of Jyväskylä Finland. World Scientific Publishing Co.
- [29] MÄKELÄ, M.M., NEITTAANMÄKI, P., (1982) *Nonsmooth optimization: Analysis and algorithms with applications to optimal control*. Singapore: World Scientific Publishing Co.
- [30] MARIO T., ALFREDO C., (2011). *A feasible directions method for nonsmooth convex optimization*. Springer-Verlag janeiro 2011. DOI 10.1007/s00158-011-0634-y.

- [31] REGINA S., RAFAIL N., NERGIZ A., YALÇIN K., (2006) *On a Modified Subgradient Algorithm for Dual Problems via Sharp Augmented Lagrangian*. Journal of Global Optimization (2006)34:55-78, Springer. DOI 10.1007/s10898-005-3270-5.
- [32] MIETTINEN, M., HASLINGER, J., (1992) *Approximation of nonmonotone multivalued differential inclusions*. IMA Journal of Numerical Analysis, 15, 475?503.
- [33] ROCKAFELLAR R., ROGER, J-B., (2009). *Variational Analysis*, Springer Dordrecht Heidelberg London New York.
- [34] ROSEN, J., SUZUKI S., (1965). *Constructing of non-linear programming test problems*. Communications of the ACM, 8, 113.
- [35] SÁNCHEZ J., (2018) *Una penalización de los multiplicadores de Lagrange*, Condiciones de Karush-Kuhn-Tucker en Programación no lineal. Departamento de Análisis Matemático.
- [36] SCHITTKOWSKI K., (2009). *Test Examples for Nonlinear Programming Codes*. Department of Computer Science. University of Bayreuth.
- [37] SCHRAMM, H., (1989). *Eine Kombination von Bundle- und Trust-Region-Verfahren zur Lösung nichtdifferenzierbarer Optimierungsprobleme*. PhD thesis, Bayreuther Mathematische Schriften, No. 30, Universität Bayreuth.
- [38] SEYRANIAN, A., LUND E., OLHOFF N., (1994). *Multiple eigenvalues in structural optimization problems*. Structural Optimization, v. 8, n. 4, pp. 207?227.
- [39] STOKEY N., LUCAS R., (1989). *Recursive Methods in Economic Dynamics*. First Edition. Cambridge, Mass.: Harvard University Press.
- [40] VANDENDERGHE, L., BOYD, S., (1996). *Semidefinite Programming*. SIAM Review, v. 38, n. 1, pp. 49?95.