

Universidad Centroccidental "Lisandro Alvarado"  
Decanato de Ciencias y Tecnología  
Licenciatura en Ciencias Matemáticas



Trabajo Especial de Grado

# UNA GENERALIZACIÓN DE CASI-DISTANCIA EN PROGRAMACIÓN CONVEXA

Br. Oliver O. Pérez S.

**Tutor:** Msc. Raquel Quintana

**Área de Conocimiento:** Optimización



---

## Resumen

---

Se estudia una casi-distancia generalizada aplicada a el método de punto proximal que tiene como casos degenerados la  $\phi$ -divergencia y la casi-distancia con núcleos homogéneos de segundo orden, esta casi-distancia se obtiene de estudiar funciones de penalidad de shift constante en el espacio primal. Se presentan los resultados de convergencia de los métodos primal y dual, esta casi-distancia puede verse como la diferencia entre una sucesión Bregman y sus aproximaciones lineales, para valores específicos de sus argumentos, con lo que la sucesión generada por el método de punto proximal definida por la generalización de la casi-distancia aplicada a el dual del problema es una solución del dual optimal.

*A Licis y Elba*

---

## Agradecimientos

---

Agradezco primeramente a Dios, porque sin él nada es posible.

Agradezco a mi madre Lisis y a mi padre Orlando por sus apoyo incondicional, a mi tía Elba por ser mi segunda madre.

a mis exitosas hermanos Hoffman y Silvia por su gran ejemplo.

Agradezco a mi tutora Raquel Quintana por sus ideas y su confianza. A su hermana Clavel Quintana por sus consejos, por su dedicación y toda su colaboración.

Finalmente, agradezco a los grandes profesores que tuve a lo largo de mi carrera, por enseñarme la rigurosidad matemática.



---

## Índice general

---

<b>Introducción</b>	<b>1</b>
<b>1. Preliminares</b>	<b>3</b>
<b>2. Método de punto proximal</b>	<b>7</b>
2.1. El concepto de regularización . . . . .	7
2.2. El algoritmo de punto proximal para optimización en $\mathbb{R}^n$ . . . . .	9
2.3. Lagrangeano aumentado . . . . .	14
2.4. Distancias de Bregman . . . . .	19
2.5. $\varphi$ -divergencia . . . . .	31
<b>3. Método de punto proximal con casi-distancia generalizada</b>	<b>45</b>
3.1. Generando la casi distancia . . . . .	47
3.1.1. Funciones de penalidad . . . . .	48
3.1.2. Shift en función penalidad . . . . .	50
3.2. La generalización de la casi-distancia . . . . .	53
3.2.1. Interpretación geométrica . . . . .	59

3.3. Métodos y teoremas . . . . .	61
3.3.1. Método proximal . . . . .	62
3.3.2. Algoritmo lagrangeano aumentado . . . . .	63
3.3.3. Resultados de convergencia . . . . .	64
<b>Conclusión</b>	<b>74</b>

---

## Introducción

---

En el marco de la programación convexa, dado el problema

$$\hat{f} = \inf\{f_0(x) : f_i(x) \leq 0 \quad i = 1, \dots, m\}, \quad (\text{P})$$

con  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ , para  $i = 0, 1, \dots, m$  donde las funciones son convexas, propias y cerradas, el problema (P) ha sido atacado y resuelto, tanto por medio de métodos teóricos como métodos computacionales, debido a que múltiples problemas de ingeniería, en sus diversas áreas, pueden ser modelados bajo este formato. Asociado al problema (P) se tiene el problema dual (D), el cual viene dado por:

$$\hat{d} = \sup\{d(\mu) : \mu \geq 0\} = \inf\{-d(\mu) : \mu \geq 0\} \quad (\text{D})$$

donde  $d(\mu) = \inf\{l(x, \mu) : x \in \mathbb{R}^n\}$  y

$$l(x, \mu) = f_0(x) + \sum_{i=1}^m \mu_i f_i(x)$$

es la función Lagrangeana.

Una forma de resolver el problema (P) consiste en utilizar los métodos de Lagrangeano aumentado, los cuales, debido a su eficiente aplicabilidad computacional en una

diversidad de problemas y su conocida relación teórica con los métodos de punto proximal, vía dualidad de Fenchel, han sido estudiados ampliamente en los últimos años. Los cuales han sido estudiados básicamente desde dos puntos de vista, el punto de vista primal-dual [3] y [13] y el punto de vista dual-primal, [2] y [4].

En los estudios que se conocen desde el punto de vista dual-primal, se utiliza el método de punto proximal para resolver el problema dual regularizando la función dual mediante casi-distancias que actúan como núcleos. En [9] se expresa que las casi-distancias asociadas a  $\varphi$  – *divergencias*, núcleos cuadráticos y distancias de Bregman todas son casos particulares de una casi-distancia generalizada. En el siguiente trabajo se encuentra un estudio exhaustivo de este artículo, el cual permite generalizar la casi-distancia entrópica Kullback-Leibler, estudiando una familia de funciones, que no necesariamente pasa a través del origen con pendiente 1.

En el Capítulo 1 se desarrollarán algunos conceptos y teoremas básicos del análisis convexo y programación no lineal.

En el Capítulo 2 se estudiará la relación entre los métodos de lagrangeano aumentado y los métodos de punto proximal.

Finalmente, en el Capítulo 3 se presentan los aportes este trabajo, ofreciendo un estudio detallado del trabajo presentado por Campos, Castillo y Hernández, [5].

# CAPÍTULO 1

---

## Preliminares

---

En este capítulo se presentaran las definiciones y los conceptos necesarios a tener en cuenta en el siguiente trabajo además de presentar y hacer referencia a los teoremas utilizados en las principales demostraciones en este trabajo.

**Definición 1.1** (ver [10] Definición A.1.1.1 )

El conjunto  $C \subset \mathbb{R}^n$  es denominado **convexo** si  $\alpha x + (1 - \alpha)y$  esta en  $C$  siempre que  $x, y$  estén en  $C$  y  $\alpha \in [0, 1]$ .

**Definición 1.2** (ver [10] Definición B.1.1.1 )

Sea  $C$  un conjunto convexo no vacío en  $\mathbb{R}^n$ . Una función  $f : C \rightarrow \mathbb{R}$  es denominada **convexa** en  $C$  cuando para todo par ordenado  $(x, y) \in C \times C$  y para todo  $\alpha \in [0, 1]$  se tiene:

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

**Definición 1.3** (ver [11] Definición 4.5)

Una función  $g : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  es **coerciva** si y sólo si

$$\lim_{\|x\| \rightarrow \infty} g(x) = +\infty.$$

**Teorema 1.1** (ver [11] Teorema 4.18)

(**Condición suficiente de optimalidad si  $f$  es de clase  $C^2$** ) Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $C^2$  en  $\mathbb{R}^n$ ,  $\nabla f(x^*) = 0^n$  y  $\nabla^2 f(x^*)$  es semidefinida positiva, entonces  $f(x^*)$  es un mínimo local estricto sobre  $\mathbb{R}$ .

**Teorema 1.2** (ver [11] Teorema 4.3)

(**Teorema fundamental de la optimalidad global**) Si  $S$  es convexa y  $f$  es convexa sobre  $S$  entonces cada mínimo local de  $f$  es un mínimo global.

**Proposición 1.1** Sea  $C$  un subconjunto convexo de  $\mathbb{R}^n$  y sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función dos veces continuamente diferenciable sobre  $\mathbb{R}^n$ .

- a) Si  $\nabla^2 f(x)$  es semidefinida positiva para todo  $x \in C$  entonces  $f$  es convexa sobre  $C$ .
- b) Si  $\nabla^2 f(x)$  es definida positiva para todo  $x \in C$  entonces  $f$  es estrictamente convexa sobre  $C$ .
- c) Si  $C$  es abierta y  $f$  es convexa sobre  $C$ , entonces  $\nabla^2 f(x)$  es semidefinida positiva para todo  $x \in C$ .

**Proposición 1.2** (ver [3] Proposición 1.2.2)

Sea  $f : \mathbb{R}^n \rightarrow [-\infty, \infty]$  una función entonces las siguientes condiciones son equivalentes:

- i) El conjunto de nivel  $\{x : f(x) \leq \lambda\}$  es cerrado para cada escalar  $\lambda$ .
- ii)  $f$  es semicontinua inferiormente sobre  $\mathbb{R}$ .
- iii) el epigrafo de  $f$  es cerrado.

**Definición 1.4** (ver [3])

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función convexa, se dirá que el vector  $d \in \mathbb{R}^n$  es un subgradiente de  $f$  en el punto  $x \in \mathbb{R}^n$

$$f(z) \geq f(x) + \langle z - x, d \rangle \quad \text{para todo } z \in \mathbb{R}^n.$$

**Definición 1.5** (ver [3])

El conjunto de todos los subgradientes de  $f$  en  $x \in \mathbb{R}^n$  es denominado subdiferencial de  $f$  en  $x$  y es denotado por  $\partial f(x)$ .

**Definición 1.6** (ver [10] Definición E.1.1.1 )

Sea  $f > -\infty$  y existe un minimizador de  $f$ . La conjugada de la función  $f$  es la función  $f^*$  definida por

$$\mathbb{R}^n \ni s \rightarrow f^* := \sup\{\langle s, x \rangle - f(x) : x \in \text{dom } f\}.$$

**Corolario 1.1** (ver [12] Corolario 23.5.1 )

Si  $f$  es una función cerrada propia convexa,  $\partial f^*$  es la inversa de  $\partial f$  en el sentido de la función multivalor, es decir  $x \in \partial f^*(x^*)$  si y solo si  $x^* \in \partial f$ .

**Proposición 1.3** (ver [10] Proposición E.1.3.1 )

Sea  $f, f_j > -\infty$  y existe un minimizador de  $f, f_j$  en  $\mathbb{R}^n$  entonces se cumple

- i) La conjugada de la función  $g(x) := f(x) + r$  es  $g^*(s) = f^*(s) - r$ .
- ii) Si  $t > 0$  la conjugada de la función  $g(x) := tf(x)$  es  $g^*(s) = tf^*(s/t)$ .
- iii) Si  $t \neq 0$  la conjugada de la función  $g(x) := f(tx)$  es  $g^*(s) = f^*(s/t)$ .
- iv) Mas generalmente: si  $A$  es un operador lineal invertible  $(f \circ A)^* = f^* \circ (A^{-1})^*$ .
- v) La conjugada de la función  $g(x) = f(x - x_0)$  es  $g^*(s) = f^*(s) + \langle s, x_0 \rangle$ .
- vi) La conjugada de la función  $g(x) := f(x) + \langle s^0, x \rangle$  es  $g^*(s) = f^*(s - s^0)$ .
- vii) Si  $f_1 \leq f_2$  entonces  $f_1^* \geq f_2^*$ .
- viii) Convexidad de la conjugación: si  $\text{dom } f_1 \cap \text{dom } f_2 \neq \emptyset$  y  $\alpha \in [0, 1]$

$$[\alpha f_1 + (1 - \alpha) f_2]^* \leq \alpha f_1^* + (1 - \alpha) f_2^*.$$

ix) La transformación Legendre-Fenchel preserva la descomposición como

$$\mathbb{R}^n := \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m} \ni x \rightarrow f(x) := \sum_{j=1}^m f_j(x_j).$$

y asumamos que  $\mathbb{R}^n$  tiene el producto escalar de un espacio de producto, se tiene

$$f^*(s_1, \dots, s_m) = \sum_{j=1}^m f^*(s_j).$$

**Teorema 1.3** (ver [10] Teorema E.1.4.1 )

Sea  $f > -\infty$  y existe un minimizador de  $f$  en  $\mathbb{R}^n$ ,  $s \in \partial f(x)$  si y solo si

$$f^*(s) + f(x) - \langle s, x \rangle = 0 \quad (o \leq 0).$$

**Teorema 1.4** (ver [1] Teorema 5.11)

(**Teorema del Valor Medio T.V.M**) Sea  $f$  una función con derivadas (finitas o infinitas) en cada uno de los puntos de un intervalo abierto  $(a,b)$  y supongamos además que  $f$  es continua en los extremos  $a$  y  $b$ . Entonces existe un punto  $c$  de  $(a,b)$  tal que

$$f(b) - f(a) = f'(c)(b - a).$$

## CAPÍTULO 2

---

### Método de punto proximal

---

En el siguiente capítulo se exponen las ideas principales que generar el método de punto proximal. Conocer las ideas que generan este método es de mucha importancia ya que es el método sobre el cual estaremos trabajando, también se estudiarán algunas variantes usando unas casi-distancias como distancia de Bregman y  $\varphi$ -divergencia, el estudio del método con estas variantes es de importancia en este trabajo ya que en el capítulo 3 se presentará una casi-distancia generalizada con la cual se aplicará el método de punto proximal que tiene como casos particulares el método de punto proximal con las casi-distancias mencionadas y estudiadas en este capítulo.

### **2.1. El concepto de regularización**

El método sobre el que estaremos trabajando es el método punto proximal, y este tiene sus orígenes en la idea de regularización de problemas mal planteados de aquí la gran importancia del estudio de este tema.

La idea de regularización surgió en conexión con problemas mal planteados. Dado un problema de la forma

$$L(f) = 0. \tag{2.1}$$

Donde  $f$  es un elemento de un conjunto  $X$  (usualmente un espacio de funciones) y  $L : X \rightarrow X$  es un operador (usualmente diferencial o integro diferencial). El problema anterior (2.1) se dice que es mal planteado cuando este no tiene solución o tiene infinitas soluciones. La idea es tratar de encontrar un nuevo operador  $L + \lambda M$  (con  $\lambda \in \mathbb{R}$ ,  $M : X \rightarrow X$ ) donde  $M$  es tal que el operador encontrado tiene una única solución .

$$L(f) + \lambda M(f) = (L + \lambda M)(f) = 0. \tag{2.2}$$

Este operador (2.2) anteriormente encontrado es bien planteado (es decir no es mal planteado) para cualquier  $\lambda > 0$  en (2.2) tiene una única solución  $f_\lambda$ , y uno espera que a medida que  $\lambda$  se acerca a 0 esta  $f_\lambda$  se aproxima a la solución del problema inicial (2.1).

El concepto de regularización, aplicado a problemas de optimización, tomando  $X = \mathbb{R}^n$  y  $L = \nabla f$ , donde  $f$  es una función convexa ( $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ) en cuyo caso la primera ecuación (2.1) se convierte en

$$\nabla f(x) = 0. \tag{2.3}$$

O equivalentemente

$$\min_{x \in \mathbb{R}^n} f(x), \tag{2.4}$$

Asumiendo que  $f$  está acotada por debajo y tomado una función  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  estrictamente convexa y coerciva, el problema no regularizado puede no tener solución pero

el problema regularizado

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda g(x), \quad (2.5)$$

tiene una única solución para cada  $\lambda > 0$  ya que  $f(x) + \lambda g$  es coerciva, además usando el hecho de que  $f$  es acotada por debajo, lo cual reduce el problema a conjuntos compactos, garantizando la existencia de soluciones, la convexidad estricta garantiza la unicidad de la solución de (2.5) teniendo una única solución de  $x(\lambda)$  y bajo algunas hipótesis razonables (incluyendo la existencia de soluciones de (2.4) puede ser probado que el  $\lim_{\lambda \rightarrow 0} x(\lambda)$  existe y resuelve (2.4). El problema con este enfoque de regularización es que, a pensar de que  $f(x) + \lambda g(x)$  es estrictamente convexa y coerciva para cada  $\lambda > 0$ , sin embargo cuando  $\lambda$  es muy pequeño esta función podría ser numéricamente casi tan mal comportado como  $f$ . Si el sistema  $\nabla f(x)$  esta mal condicionado entonces el sistema  $(\nabla f(x) + \nabla g(x))$  sera también mal condicionado cuando  $\lambda$  se aproxima a 0. A pesar del hecho de que esta tiene una solución para todo  $\lambda > 0$ .

## 2.2. El algoritmo de punto proximal para optimización en $\mathbb{R}^n$ .

Presentaremos ahora el método de punto proximal en su forma clásica, principal método de estudio en este trabajo y en el cual aplicaremos la casi-distancia presentada en el capítulo 3, que es nuestro principal tema de estudio.

Como ya observamos en la sección anterior el solo hecho de regularización no es suficiente para lograr tener un problema bien comportado. Con el objetivo de superar las dificultades mencionadas en la regularización es deseable desarrollar un enfoque de regularización que no requiera la regularización del parámetro  $\lambda$  tendiendo a cero

(es decir trabajar con una constante  $\lambda$ ). El algoritmo de punto proximal alcanza tal objetivo. Este genera una sucesión  $\{x^{k+1}\} \subset \mathbb{R}^n$  de la siguiente forma.

$$x^0 \in \mathbb{R}^n, \quad (2.6)$$

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \|x - x^k\|^2\}. \quad (2.7)$$

Donde cada  $\lambda_k$  es un número real satisfaciendo  $0 < \lambda_k < \lambda$ . Para algún  $\lambda > 0$  (que incluye el caso  $\lambda_k$  constante) y  $\|\cdot\|$  es la norma euclidiana.

Mostraremos que bajo algunas hipótesis razonables la sucesión generadas por (2.6) y (2.7) convergen a un minimizador de  $f$ . El enfoque que se sigue para probar esta convergencia pasa a través del concepto de no expansividad y trabaja en hechos mucho más generales que (2.3) la idea es seguir un enfoque de Fejer convergencia.

**Definición 2.1** (ver [6] Sección 2)

Una sucesión se dice que es **Féjer Convergente** en un conjunto  $U \in \mathbb{R}^n$  con respecto a la norma euclidiana si  $\|y^{k+1} - u\| \leq \|y^k - u\|$  para todo  $k \geq 0$  y para todo  $u \in U$ .

Se tiene el siguiente resultado.

**Proposición 2.1** (ver [6] Proposición 2.1)

Si  $\{y^k\}$  es Féjer convergente para  $U \neq \emptyset$  entonces  $\{y^k\}$  es acotada. Si un punto de acumulación  $y$  de  $\{y^k\}$  pertenece a  $U$ , entonces  $y = \lim_{k \rightarrow \infty} y^k$ .

**Demostración:**

Para el primer condicional, por ser  $\{y^k\}$  Féjer convergencia tenemos que  $\|y^k - u\| \leq \|y^0 - u\|$ , para cada  $u \in U$ , con lo que la sucesión  $\{y^k\}$  esta contenida en la bola de centro  $u$  y radio  $\|y^0 - u\|$ , con esto se tiene que la sucesión es acotada.

Para el segundo condicional, sea  $\{y^{j_p}\}$  una subsucesión de  $\{y^k\}$  tal que  $\lim_{k \rightarrow \infty} y^{j_p} = y$ , por definición de límite se tiene que para todo  $\epsilon > 0$  existe un  $n \in \mathbb{N}$

tal que si  $p \geq n$  entonces  $\|y^{jp} - y\| \leq \epsilon$ , y por definición de subsucesión, existe un  $m \in \mathbb{N}$  tal que  $y^m = y^{jp}$ , así que  $\|y^m - y\| < \epsilon$ , por ser  $\{y^k\}$  Féjer convergente tenemos que si  $k \geq m$  entonces  $\|y^k - y\| < \epsilon$  y esto no es mas que la definición de límite con lo que tenemos  $\lim_{k \rightarrow \infty} y^k = y$ .

Con el teorema anterior ya se puede probar la convergencia del algoritmo de punto proximal.

**Teorema 2.1** (ver [6] Teorema 2.1)

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  convexa y continuamente diferenciable. Supongamos que el conjunto  $U$  de minimizadores de  $f$  en  $\mathbb{R}^n$  es no vacío. Entonces la sucesión generada por el método de punto proximal converge a un punto  $x^* \in U$ .

**Demostración:**

Se divide la prueba en cuatro pasos:

**Paso 1** Mostraremos que  $\{y^k\}$  esta bien definida, la buena definición de la sucesión pasa por que la función  $f_k(x)$  dada por  $f_k(x) = f(x) + \lambda \|x - x^k\|^2$ , siempre tenga mínimo y la obtención de este mínimo lo tenemos garantizando ya que  $f_k(x)$  es continua, convexa, acotada y que  $\lim_{\|x\| \rightarrow \infty} f_k(x) = +\infty$ .

Observemos que  $f_k(x)$  es continua, sabemos que  $\|\cdot\|$  es continua además la función  $g(x) = x^2$  es continua como la composición de funciones continuas es continua teniendo que  $\|x - x^k\|^2$  es continua,  $f(x)$  es continua por hipótesis ya que  $f$  es continuamente diferenciable, además como la suma de funciones continuas es continua se tiene que  $f(x) + \|x - x^k\|^2$  es continua así que  $f_k(x)$  es continua. Con respecto a la convexidad se tiene que  $f(x)$  es convexa por hipótesis y  $\|x - x^k\|^2$  es convexa ya que la función  $g(x) = x^2$  lo es y como la suma de funciones convexa es convexa tenemos que:  $f(x) + \|x - x^k\|^2$  es convexa con lo que  $f_k(x)$  es convexa. Por hipótesis tiene que  $f(x)$  es acotada inferiormente y sabemos que la función  $g(x) = x^2$  es acotada inferiormente

por cero, con lo que si  $u \in U$  se tiene que  $u \leq f(x) + \|x - x^k\|^2$  con esto  $f_k(x)$  es acotada inferiormente. Y por último el  $\lim_{\|x\| \rightarrow \infty} f_k(x) = +\infty$  se satisface ya que  $\|x - x^k\|^2$  es coerciva, con lo que  $f_k(x)$  es coerciva.

Con todas estas condiciones solo basta evaluar  $f_k(x)$  sobre un conjunto compacto para obtener el mínimo el cual es único.

**Paso 2** Comprobemos que  $\{x^k\}$  es Féjer convergente para  $u \in U$ , para esto solo basta probar que:  $\|x^{k+1} - u\|^2 \leq \|x^k - u\|^2 - \|x^{k+1} - x^k\|^2$  para toda  $k \geq 0$  y todo  $u \in U$  sumando la siguiente inecuación  $0 \leq \|x^{k+1} - x^k\|^2$  nos queda que  $\|x^{k+1} - u\|^2 \leq \|x^k - u\|^2$  y esto implica que  $\|x^{k+1} - u\| \leq \|x^k - u\|$ .

Probando la desigualdad:

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

Se tiene que

$$\begin{aligned} \|x^{k+1} - u\|^2 &= \|x^k - x^{k+1} + x^{k+1} - u\|^2, \\ &= \|x^k - x^{k+1}\|^2 + \|x^{k+1} - u\|^2 + 2\langle x^k - x^{k+1}, x^{k+1} - u \rangle. \end{aligned} \quad (2.8)$$

Ya que,

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \|x - x^k\|^2\}, \\ &= \operatorname{argmin}_{x \in \mathbb{R}^n} \{f_k(x)\}. \end{aligned}$$

Derivando  $f_k(x)$  ya que es diferenciable y como  $x^{k+1}$  es el mínimo de  $f_k(x)$ .

$$\begin{aligned} 0 &= \nabla f_k(x^{k+1}), \\ &= \nabla f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k). \end{aligned} \quad (2.9)$$

Despejando de (2.9) tenemos

$$\frac{\nabla f(x^{k+1})}{2\lambda_k} = (x^k - x^{k+1}). \quad (2.10)$$

Sustituyendo (2.8) en (2.10) tenemos que

$$\begin{aligned}
\|x^k - u\|^2 - \|x^k - x^{k+1}\|^2 - \|x^k - u\|^2 &= 2 \left\langle \frac{\nabla f(x^{k+1})}{2\lambda_k}, x^{k+1} - u \right\rangle, \\
&= \frac{1}{\lambda_k} \langle \nabla f(x^{k+1}), x^{k+1} - u \rangle, \\
&= \frac{1}{\lambda_k} [f(x^{k+1}) - f(u)]. \quad (2.11)
\end{aligned}$$

Obtenemos de (2.11) y del hecho de ser  $f(x)$  diferenciable y como  $u$  es un minimizador de  $f(x)$  tenemos que  $f(x^{k+1}) \geq f(u)$  con lo que  $f(x^{k+1}) - f(u) \geq 0$  así que  $\frac{1}{\lambda_k} [f(x^{k+1}) - f(x)] \geq 0$  con lo que de (2.11) se obtiene.

$$\begin{aligned}
\|x^k - u\|^2 - \|x^k - x^{k+1}\|^2 - \|x^k - u\|^2 &\geq 0, \\
\|x^k - u\|^2 - \|x^{k+1} - x^k\|^2 &\geq \|x^{k+1} - u\|^2. \quad (2.12)
\end{aligned}$$

Con esto se puede decir que  $\{x^k\}$  es Féjer convergente.

**Paso 3** Probaremos que el  $\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0$ .

Ya que  $\{\|x^k - u\|\}$  es decreciente, no negativa y acotado por cero podemos decir que  $\{\|x^k - u\|\}$  converge a cero y además como  $\{x^k\}$  es Féjer convergente se tiene que  $\{\|x^{k+1} - u\|\}$  también converge a cero.

$$\begin{aligned}
\|x^{k+1} - x^k\|^2 &\leq \|x^k - u\|^2 - \|x^{k+1} - u\|^2, \text{ por (2.12)} \\
\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\|^2 &\leq \lim_{k \rightarrow \infty} \|x^k - u\|^2 - \lim_{k \rightarrow \infty} \|x^{k+1} - u\|^2, \\
&= \left( \lim_{k \rightarrow \infty} \|x^k - u\| \right)^2 - \left( \lim_{k \rightarrow \infty} \|x^{k+1} - u\| \right)^2, \\
&= 0 + 0; \text{ por conv. de } \{\|x^k - u\|\} \text{ y } \{\|x^{k+1} - u\|\}, \\
&= 0. \quad (2.13)
\end{aligned}$$

Además como  $\|x^{k+1} - x^k\|^2 \geq 0$ , tenemos que  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\|^2 \geq 0$  y con (2.13)

$$0 \leq \lim_{k \rightarrow \infty} \|x^{k+1} - x^k\|^2 \leq 0. \quad (2.14)$$

Con (2.14) podemos decir que  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\|^2 = 0$  y por continuidad de  $g(x) = x^2$  y de la norma  $\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0$ .

**Paso 4** Debemos probar que  $\{x^k\}$  tiene puntos de acumulación, todos pertenecen a  $U$ .

Sea  $u$  un punto de acumulación de  $\{x^k\}$  y  $\{x^{j_k}\}$  una subsucesión de  $\{x^k\}$  tal que  $\lim_{k \rightarrow \infty} x^{j_k} = u$ .

Ya que  $x^{j_k+1}$  es el minimizador de  $f_k(x)$  tenemos que:

$$\nabla f(x^{j_k+1}) = 2\lambda_k(x^{j_k} - x^{j_k+1}). \quad (2.15)$$

Del paso 3 se tiene que  $\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0$  con esto se tiene que  $\lim_{k \rightarrow \infty} (x^{j_k+1} - x^{j_k}) = 0$  y esto implica que  $\lim_{k \rightarrow \infty} x^{j_k+1} = \lim_{k \rightarrow \infty} x^{j_k}$

Así que

$$\lim_{k \rightarrow \infty} x^{j_k+1} = u.$$

Además como  $\lambda_k \leq \bar{\lambda}$  y como  $f$  es continuamente diferenciable tenemos  $\nabla f(u) = 0$ , por la convexidad de  $f$ ,  $u \in U$ , gracias al teorema fundamental de optimalidad global.

Con el paso 2 y 4 mas la proposición tenemos que existe un  $x^* \in U$  tal que  $x^* = \lim_{k \rightarrow \infty} x^k$ .

### 2.3. Lagrangeano aumentado

Discutiremos ahora la utilidad del algoritmo de punto proximal y su utilidad en optimización. Lo primero que debemos tener en cuenta es que el método de punto proximal puede ser visto como un esquema conceptual mas que como un algoritmo implementable. Aunque en general  $f_k(x) = f(x) + \lambda_k \|x^k - x\|^2$  es mas fácil de minimizar que  $f(x)$  pero para encontrar el minimizador de  $f(x)$  se requiere minimizar  $f_k(x)$  muchas ve-

ces. Así que este método es bueno solo cuando  $f_k(x)$  es mas fácil de minimizar que  $f(x)$ .

En esta sección el método de Lagrangeano aumentado el cual sustituye un problema original restringido por una sucesión de problemas sin restricción, mostraremos que este método es justamente un caso particular del método de punto proximal.

El problema a considerar es:

$$\begin{aligned} & \text{mín } f(x) \\ & \text{s.a } g_i(x) \leq 0 \quad (1 \leq i \leq m). \end{aligned} \tag{2.16}$$

Con  $f, g_i : \mathbb{R}^n \rightarrow \mathbb{R}$  convexas y diferenciables.

El Lagrangeano estándar para este problema anterior es  $L(x, y) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  dada por

$$L(x, y) = \begin{cases} f(x) + \sum_{i=1}^m y_i g_i(x) & \text{si } y \geq 0 \\ +\infty & \text{otro caso,} \end{cases} \tag{2.17}$$

Una solución para el problema (2.16) es generar dos sucesiones  $\{x^k\}$  y  $\{y^k\}$  tal que dado  $y^k \geq 0$ ,  $x^k = \underset{x \in \mathbb{R}^n}{\text{argmin}} \{L(x, y^k)\}$  actualizando a  $y$  de manera conveniente. Computacionalmente los problemas están relacionados con la discontinuidad de  $L(x, \cdot)$  surgiendo la necesidad de mejorarlos, por esto se desarrolla el método de Lagrangeano aumentado, el cual tiene mejores propiedades de convergencia.

El Lagrangeano aumentado para el problema es el siguiente, para  $x \in \mathbb{R}^n$ , sea  $x_j^+ \in \mathbb{R}^n$  definida como  $x_j^+ = \text{mín}\{x_j, 0\}$ . tomando  $\alpha \geq 0$  se define el Lagrangeano aumentado  $L_\alpha(x, y) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  como

$$L_\alpha(x, y) = f(x) + \frac{1}{4\alpha} \sum_{i=1}^m \{[(y_i + 2\alpha g_i(x))^+]^2 - y_i^2\}, \tag{2.18}$$

$L_\alpha(x, y)$  es diferenciable.

En Efecto,

$g_i(x), y_i$  y  $h(x) = x^2$  es diferenciable por regla de la cadena y ya que la suma resta y producto de funciones diferenciables es también diferenciable tenemos que  $\frac{1}{4\alpha} \sum_{i=1}^m \{[(y_i + 2\alpha g_i(x))^+]^2 - y_i^2\}$  es diferenciable y como por hipótesis  $f(x)$  es diferenciable tenemos que  $f(x) + \frac{1}{4\alpha} \sum_{i=1}^m \{[(y_i + 2\alpha g_i(x))^+]^2 - y_i^2\}$  con lo que  $L_\alpha(x, y)$  es una vez diferenciable, la doble diferenciability depende de si  $f(x), g_i(x)$  con  $i = 1, \dots, m$  es dos veces diferenciables, con derivada.

$$\nabla_x L_\alpha(x, y) = \nabla f(x) + \frac{1}{4\alpha} \sum_{i=1}^m 2(y_i + 2\alpha g_i(x))^+ 2\alpha \nabla g_i(x), \quad (2.19)$$

$$= \nabla f(x) + \sum_{i=1}^m (y_i + 2\alpha g_i(x))^+ \nabla g_i(x). \quad (2.20)$$

El método de lagrangeano aumentado genera  $\{x^k\}$  y  $\{y^k\}$  a través de

$$y^0 \in \mathbb{R}_+^n, \quad (2.21)$$

dado  $y^k \in \mathbb{R}^m$

$$x^k = \operatorname{argmin}_{x \in \mathbb{R}^n} \{L_\alpha(x, y^k)\}, \quad (2.22)$$

$$y_i^{k+1} = (y_i^k + 2\alpha g_i(x^k))^+. \quad (2.23)$$

Consideremos ahora la función objetivo dual  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  definida por

$$\varphi(y) = \min_{x \in \mathbb{R}^n} \{L(x, y)\} \quad (2.24)$$

Como  $L$  es convexa  $\varphi$  también lo es. Ahora en el siguiente teorema se mostrara que las sucesiones por (2.21)-(2.23) es la misma que la generada por el algoritmo de punto proximal aplicada a  $-\varphi$ .

**Teorema 2.2** (ver [6] Teorema 7.1)

Sea  $\{\bar{y}^k\}$  la sucesión generada por el método de punto proximal para  $\min\{-\varphi(y) : y \in \mathbb{R}^m\}$  con  $\lambda_k = \frac{1}{4\alpha}$ , y  $\{y^k\}$  la sucesión generada por (2.33)-(2.35). Si  $\bar{y}^0 = y^0$  entonces  $\bar{y}^k = y^k$

**Demostración:**

Se procede por inducción.

Para demostrar que para  $k = 1$  se tiene que  $\bar{y}^1 = y^1$ , teniendo que  $\bar{y}^0 = y^0$  el razonamiento es igual a demostrar que  $\bar{y}^{k+1} = y^{k+1}$  teniendo que  $\bar{y}^k = y^k$ , con lo que solo haremos la prueba para el ultimo caso.

Suponga que  $\bar{y}^n = y^n$  ya que.

$$\bar{y}^{k+1} = \operatorname{argmin}_{y \in \mathbb{R}^n} \left\{ -\varphi(y) + \frac{1}{4\alpha} \|y - y^k\|^2 \right\}. \quad (2.25)$$

Como

$$\begin{aligned} 0 &= \partial\varphi_k(y^{k+1}) = \partial(-\varphi(\bar{y}^{k+1}) + \frac{1}{4\alpha} \|\bar{y}^{k+1} - y^k\|^2), \\ &= -\partial\varphi(\bar{y}^{k+1}) + \frac{1}{2\alpha}(\bar{y}^{k+1} - y^k), \end{aligned} \quad (2.26)$$

$$\begin{aligned} &\Rightarrow \frac{1}{2\alpha}(\bar{y}^{k+1} - y^k) = \partial\varphi(\bar{y}^{k+1}), \\ &\Rightarrow \frac{1}{2\alpha}(\bar{y}^{k+1} - y^k) \in \partial\varphi(\bar{y}^{k+1}). \end{aligned} \quad (2.27)$$

Por definición de  $\partial f$  y  $\varphi$ , (2.27) es equivalente a

$$\begin{aligned} &\frac{1}{2\alpha} \langle \bar{y}^{k+1} - y^k, \bar{y}^{k+1} - u \rangle \leq \varphi(\bar{y}^{k+1}) - \varphi(u), \\ \varphi(u) &\leq \varphi(\bar{y}^{k+1}) - \frac{1}{2\alpha} \langle \bar{y}^{k+1} - y^k, \bar{y}^{k+1} - u \rangle \text{ Para todo } u \geq 0. \end{aligned} \quad (2.28)$$

Ya que (2.28) determina  $\bar{y}^{k+1}$  excepcionalmente, basta probar que (2.28) se satisface con  $y^{k+1}$  sustituyendo a  $\bar{y}^{k+1}$  por (2.20),(2.21),(2.22),(2.23)

$$\begin{aligned}
0 = \nabla_x L(x^k, y^k) &= \nabla f(x) + \sum_{i=1}^n (y_i^k + 2\alpha g_i(x^k))^+ \nabla g_i(x^k), \\
&= \nabla f(x) + \sum_{i=1}^n (y_i^{k+1} \nabla g_i(x^k)) \in \partial_x L(x^k, y^k),
\end{aligned} \tag{2.29}$$

(2.29) implica que  $x^k$  minimiza  $L(\cdot, y^{k+1})$  y por (2.24)

$$\varphi(y^{k+1}) = \nabla f(x^k) + \sum_{i=1}^n (y_i^{k+1} g_i(x^k)). \tag{2.30}$$

Por otro lado por (2.23) tenemos que

$$\begin{aligned}
y_i^{k+1} > 0 &\Rightarrow y_i^{k+1} = y_i^k + 2\alpha g_i(x^k) \Rightarrow y_i^{k+1} - y_i^k = 2\alpha g_i(x^k), \\
y_i^{k+1} = 0 &\Rightarrow y_i^{k+1} - y_i^k = -y_i^k.
\end{aligned}$$

Con lo que

$$y_i^{k+1} - y_i^k = \max\{-y_i^k, y_i^k + 2\alpha g_i(x^k)\} \geq 2\alpha g_i(x^k).$$

Además

$$\begin{aligned}
*y_i^{k+1} > 0 &\Rightarrow y_i^{k+1}(y_i^{k+1} - y_i^k) = y_i^{k+1} 2\alpha g_i(x^k), \\
y_i^{k+1} = 0 &\Rightarrow y_i^{k+1}(y_i^{k+1} - y_i^k) = 0(y_i^{k+1} - y_i^k) = 0.
\end{aligned}$$

Lo cual implica

$$y_i^{k+1}(y_i^{k+1} - y_i^k) = y_i^{k+1} 2\alpha g_i(x^k).$$

Los términos de la expresión  $\frac{1}{2\alpha} \langle \bar{y}^{n+1} - y^n, \bar{y}^{n+1} - u \rangle$  para algún  $u \geq 0$  y sustituyendo  $\bar{y}^{n+1}$  por  $y^{n+1}$  es de la forma.

$$\begin{aligned}
\frac{1}{2\alpha} \langle y^{k+1} - y^k, y^{k+1} - u \rangle &= \frac{1}{2\alpha} y^{k+1}(y^{k+1} - u) - \frac{1}{2\alpha} y^k(y^{k+1} - u), \\
&= \frac{1}{2\alpha} y_i^{k+1} 2\alpha g_i(x^k) - \frac{1}{2\alpha} u(y^{k+1} - u), \\
&\leq y_i^{k+1} g_i(x^k) - \frac{1}{2\alpha} u 2\alpha g_i(x^k), \\
&= y_i^{k+1} g_i(x^k) - u g_i(x^k).
\end{aligned} \tag{2.31}$$

De (2.31) y (2.32) se obtiene

$$\begin{aligned}
\varphi(y^{k+1}) - \frac{1}{2\alpha} \langle y^{k+1} - y^k, y^{k+1} - u \rangle &\geq f(x^k) + \sum_{i=1}^m y_i^{k+1} g_i(x^k) - \sum_{i=1}^m y_i^{k+1} g_i(x^k) + \\
&\quad \sum_{i=1}^n u g_i(x^k), \\
&= f(x^k) + \sum_{i=1}^n u g_i(x^k), \\
&= L(x^k, u), \\
&\geq \min_{x \in \mathbb{R}^n} L(x, u), \\
&= \varphi(u).
\end{aligned} \tag{2.32}$$

Con (2.32) se muestra que (2.28) se mantiene con  $y^{k+1}$  así que  $\bar{y}^{k+1} = y^{k+1}$  lo cual completa la prueba por inducción.

El teorema 2.2 combinado con el teorema 2.1 implica la convergencia de la sucesión  $\{y^k\}$  generada por el método de Lagrangeano Aumentado para un minimizador  $y^*$  de la función dual objetivo  $\varphi$ .

## 2.4. Distancias de Bregman

En esta sección se presentaremos un tipo de funciones que satisfaciendo determinadas características son denominadas funciones Bregman, utilizando este tipo de funciones se definen una casi-distancia denominada distancia de Bregman. Además de presentar en esta sección este tipo de función y sus respectivas distancia, se definen características y propiedades de las funciones Bregman y que nos permiten aplicar la distancia de Bregman en el método de punto proximal y obtener la convergencia de la sucesión generada por el método.

Sea  $S$  un subconjunto abierto y convexo de  $\mathbb{R}^n$  y  $\bar{S}$  es la clausura. Y una función real

convexa  $h$  definida en  $\bar{S}$ . Definamos la distancia de Bregman como la función denotada por  $D_h : \bar{S} \times S \rightarrow \mathbb{R}$  por:

$$D_h(x, y) = h(x) - h(y) - \nabla h(y)^t(x - y) \quad (2.33)$$

La función  $h$  es llamada una función de Bregman ( y  $D_h$  la distancia de Bregman dada por  $h$ ) si cumple las siguientes condiciones:

$B_1$ )  $h$  es continuamente diferenciable en  $S$ .

$B_2$ )  $h$  es estrictamente convexa y continua en  $\bar{S}$ .

$B_3$ ) Para todo  $\delta \in \mathbb{R}$  los siguientes conjuntos son acotados.

$$\Gamma_1(y, \delta) = \{x \in \bar{S} : D_h(x, y) \leq \delta\} \quad \text{para todo } y \in S.$$

$$\Gamma_2(x, \delta) = \{y \in S : D_h(x, y) \leq \delta\} \quad \text{para todo } x \in \bar{S}.$$

$B_4$ ) si  $\{y^k\} \subset S$  converge a  $y^*$  entonces  $D_h(y^*, y^k)$  converge a 0.

$B_5$ ) Si  $\{x^k\} \subset \bar{S}$  y  $\{y^k\} \subset S$  son sucesiones tal que  $\{x^k\}$  es acotada,  $\lim_{k \rightarrow \infty} y^k = y^*$  y  $\lim_{k \rightarrow \infty} D_h(x^k, y^k) = 0$ , entonces  $\lim_{k \rightarrow \infty} x^k = y^*$ .

**Definición 2.2** (ver [6] Sección 9)

Una función Bregman  $h$  es llamada **frontera coerciva**. Si  $\{y^k\} \subset S$  es tal que  $\lim_{k \rightarrow \infty} y^k = y$  con  $y \in \partial S$  entonces  $\lim_{k \rightarrow \infty} \nabla h(y^k)^t(x - y^k) = -\infty$  para todo  $x \in S$ .

**Definición 2.3** (ver [6] Sección 9)

Una función  $h$  es llamada **zona coerciva**. Si para cada  $y \in \mathbb{R}^n$  existe  $x \in S$  tal que  $\nabla h(x) = y$ .

**Proposición 2.2** (ver [6] Proposición 9.1)

Si  $f$  es una función Bregman con zona  $S$  entonces:

i)  $D_h(x, y) - D_h(x, z) - D_h(z, y) = \langle \nabla h(y) - \nabla h(z), z - x \rangle \quad \forall x \in \bar{S}; \forall y, z \in S.$

ii)  $\nabla_x D_h(x, y) = \nabla h(x) - \nabla h(y)^t \quad \forall x, y \in S.$

iii)  $D_h(\cdot, y)$  es estrictamente convexa  $\forall x \in S.$

***Demostración:***

*i)*

$$\begin{aligned}
D_h(x - y) - D_h(x, z) - D_h(z, y) &= \cancel{h(x)} - \cancel{h(y)} - \nabla h(y)^t(x - y) - \cancel{h(x)} \\
&\quad + \cancel{h(z)} + \nabla h(z)^t(x - z) - \cancel{h(z)} + \cancel{h(y)} \\
&\quad + \nabla h(z)^t(z - y), \\
&= \langle \nabla h(y)^t, (z - y) - (x - y) \rangle \\
&\quad - \nabla h(z)^t(z - x), \\
&= \langle \nabla h(z)^t - \nabla h(y)^t, (z - x) \rangle.
\end{aligned}$$

*ii)*

$$\begin{aligned}
\nabla[D_h(x, y)] &= \nabla[h(x) - h(y) - \nabla h(y)^t(x - y)], \\
&= \nabla h(x) - \nabla h(z)^t.
\end{aligned}$$

*iii)* Fijemos  $y \in S$  y  $x, \bar{x} \in \bar{S}$

$$\begin{aligned}
D_h(\alpha x + (1 - \alpha)\bar{x}, y) &= h(\alpha x + (1 - \alpha)\bar{x}) - h(y) - \nabla h(y)^t(\alpha x + (1 - \alpha)\bar{x} - y), \\
&< \alpha h(x) + (1 - \alpha)h(\bar{x}) - h(y) - \nabla h(y)^t(\alpha x + (1 - \alpha)\bar{x} - y), \\
&= \alpha h(x) + (1 - \alpha)h(\bar{x}) - h(y) - \nabla h(y)^t \alpha x - \nabla h(y)^t \bar{x} \\
&\quad + \nabla h(y)^t \alpha \bar{x} + \nabla h(y)^t y, \\
&= \alpha h(x) + (1 - \alpha)h(\bar{x}) - (\alpha + (1 - \alpha))h(y) - \nabla h(y)^t \alpha x \\
&\quad - \nabla h(y)^t \bar{x} + \nabla h(y)^t \alpha \bar{x} + (\alpha + (1 - \alpha))\nabla h(y)^t y, \\
&= \alpha h(x) + (1 - \alpha)h(\bar{x}) - \alpha h(y) - (1 - \alpha)h(x) - \nabla h(y)^t \alpha x \\
&\quad - \nabla h(y)^t \bar{x} + \nabla h(y)^t \alpha \bar{x} + \alpha \nabla h(y)^t y + (1 - \alpha)\nabla h(y)^t y, \\
&= \alpha[h(x) - h(y) - \nabla h(y)^t(x - y)] \\
&\quad + (1 - \alpha)[h(\bar{x}) - h(y) - \nabla h(y)^t(\bar{x} - y)], \\
&= \alpha[D_h(x, y)] + (1 - \alpha)[D_h(\bar{x} - y)].
\end{aligned}$$

Con respecto a las condiciones  $B_1 - B_5$  las primeras dos son condiciones que dependen de las características de la función  $h$  con respecto a la función  $B_3$  esta condición se obtiene gracias a la condición  $B_2$  de continuidad sabemos que continuidad implica semi-continuidad inferior y esto nos implica conjunto de nivel acotados con lo que tenemos  $B_3$  garantizado.

### Prueba

( $\Rightarrow$ )

$$\begin{aligned} D_h(x, y) = 0 &\Rightarrow h(x) - h(y) - \nabla h(t)^t(x - y) = 0, \\ &\Rightarrow h(x) - h(y) = \nabla h(t)^t(x - y). \end{aligned}$$

( $\Leftarrow$ )

$$\begin{aligned} x = y &\Rightarrow D_h(x, x) = h(x) - h(x) - \nabla h(x)^t(x - x), \\ &\Rightarrow D_h(x, x) = \cancel{h(x)} - \cancel{h(x)} - \nabla h(x)^t(x - x), \\ &\Rightarrow D_h(x, x) = 0 - \nabla h(x)^t(0), \\ &\Rightarrow D_h(x, x) = 0 \text{ ya que } \nabla h(x)^t \text{ es distinto de } \infty. \end{aligned}$$

(Probando  $B_4$ )

Como el  $\lim_{k \rightarrow \infty} x^k = x^* \Leftrightarrow \forall \varepsilon > 0 \exists N_\varepsilon \in \mathbb{N}$  tal que  $\forall n \geq N_\varepsilon \Rightarrow \|x^n - x^*\| < \varepsilon$ .

Como  $h$  es continuamente diferenciable,  $h$  es continua y  $\lim_{k \rightarrow \infty} x^k = x^*$  y entonces  $\lim_{k \rightarrow \infty} h(x^k) = h(x^*)$  con lo que por definición de límite tenemos que  $\forall \varepsilon > 0 \exists \bar{N}_\varepsilon \in \mathbb{N}$  tal que  $\forall n \geq \bar{N}_\varepsilon \Rightarrow \|x^n - x^*\| < \varepsilon$ .

Por otro lado como  $\nabla h(x)$  es continua y  $\{x^k\}$  es una sucesión convergente  $\nabla h(x)$  también lo es por continuidad y como toda sucesión convergente es acotada existe,  $M \in \mathbb{R}$  tal que  $\|\nabla h(x^k)\| < M \forall k \in \mathbb{N}$ .

Sea  $N^* = \max\{N_\varepsilon, \bar{N}_\varepsilon\}$  como  $D_h(x^*, x^k) = h(x^*) - h(x^k) - \nabla h(x^k)^t(x^* - x^k)$  tenemos:

$$\begin{aligned} \|D_h(x^*, x^k)\| &= \|h(x^*) - h(x^k) - \nabla h(x^k)^t(x^* - x^k)\| . \\ &\leq \|h(x^*) - h(x^k)\| + \|\nabla h(x^k)^t(x^* - x^k)\| . \\ &\leq \|h(x^*) - h(x^k)\| + \|\nabla h(x^k)^t\| \|x^* - x^k\| . \\ &\leq \varepsilon + M\varepsilon . \\ &= (1 + M)\varepsilon \text{ para } k > N^* . \end{aligned}$$

Con esto tenemos que  $\lim_{k \rightarrow \infty} D_h(x^*, x^k) = 0$ .

Ahora algunos ejemplos de funciones Bregman:

**Ejemplo 2.1**  $S = \mathbb{R}^n$ ,  $h(x) = x^t M x$ , (ver Figura 2.1 para  $m=1$ ), con  $M \in \mathbb{M}_{n \times n}$  simétrica y definida positiva. En este caso  $D_h(x, y) = (x - y)^t M (x - y) = \|x - y\|^2$ .

$$\begin{aligned} h(x) &= x^t M x \\ &= \sum_{i,j=1}^n h_{i,j} x_i x_j \text{ por ser una forma cuadrática.} \end{aligned}$$

Además para cada  $i, j = \{1, \dots, m\}$   $h_{i,j} x_i x_j$  son funciones cuadráticas y como cada  $h_{i,j} \geq 0$ ,  $h_{i,j} x_i x_j$  es estrictamente convexa, continuas y continuamente diferenciables.

Por otro lado

$$\begin{aligned}
D_h(x, y) &= x^t M x - y^t M y - \nabla(y^t M y)^t(x - y), \\
&= x^t M x - y^t M y - (M y + y^t M)^t(x - y), \\
&= x^t M x - y^t M y - (M y + (M^t y)^t)(x - y), \\
&= x^t M x - y^t M y - (M y + (M y)^t)^t(x - y), \\
&= x^t M x - y^t M y - (2M y)^t(x - y), \\
&= x^t M x - y^t M y - (2y^t M)(x - y), \\
&= x^t M x - y^t M y - 2y^t M x + 2y^t M y, \\
&= (x^t - y^t)M x - y^t(x - y), \\
&= (x^t - y^t)M x - (x - y)^t M y, \\
&= (x^t - y^t)M x - (x^t - y^t)M y, \\
&= (x^t - y^t)(M x - M y), \\
&= (x^t - y^t)M(x - y). \text{(ver Figura 2.2 para } m=1\text{)}
\end{aligned}$$

Observemos que  $h(x)$  es frontera coerciva y zona coerciva.

$\partial S = \{(x_1 \cdots x_n) : x_i = \pm\infty, \text{ con } 1 \leq i \leq n\}$ . Si  $x_j = +\infty$  y  $\{y^k\}$  para algún  $1 \leq j \leq n$ , en la sucesión  $\{y^k\}$  se tiene que

$$\begin{aligned}
\lim_{k \rightarrow +\infty} \nabla h(y^k)^t(x - y) &= \lim_{k \rightarrow +\infty} 2(y^k)^t M(x - y^k). \\
&= -\infty.
\end{aligned}$$

Con lo que se cumple que es frontera coerciva.

Como  $\nabla h(x) = 2Mx$  se tiene que para cada  $y \in \mathbb{R}^n$  existe  $x$  tal que  $2Mx = y$  con lo que se cumple que es zona coerciva.

**Ejemplo 2.2**  $S = \mathbb{R}_{++}^n$ ,  $h(x) = \sum_{j=1}^n x_j \log x_j$  (ver Figura 2.4 para  $n=1$ ), extendiendo la continuidad a  $\partial \mathbb{R}_+^n$  con la condición que  $0 \log 0 = 0$  en esta caso  $D_h(x, y) =$

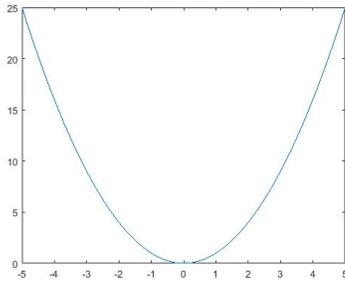


Figura 2.1:  $h(x) = x^2$ .

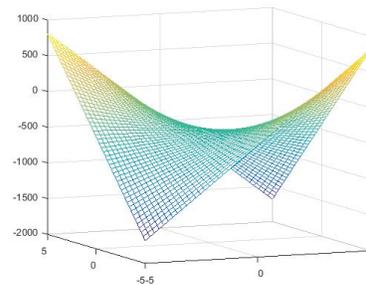


Figura 2.2:  $D_h(x, y) = (x - y)^2$ .

$\sum_{j=1}^n (x_j \log \frac{x_j}{y_j} + y_j - x_j)$  ver Figura 2.5 para  $n=1$ , la cual es la divergencia de Kullback-Leibler.

Tenemos que  $x_j$  y  $\log x_j$  son continuamente diferenciables, con lo que  $\sum_{j=1}^n x_j \log x_j$  es continuamente diferenciable. En la siguiente gráfica se observamos que la función  $f(x) = x \log x$  para  $x > 0$  es una función convexa.

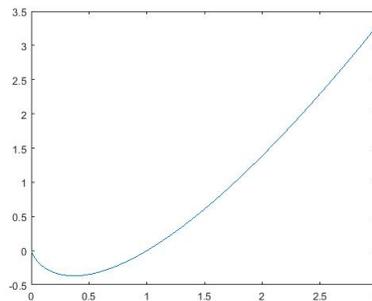


Figura 2.3:  $h(x) = x \log(x)$

La convexidad de  $h$  se obtiene de que  $h(x) = \sum_{j=1}^n x_j \log x_j$  (ver Figura 2.3) es suma de  $n$ -veces la gráfica anterior con lo que  $h(x)$  es convexa ya que la suma de funciones convexas preserva la convexidad.

Por otro lado:

$$\begin{aligned}
D_x(x, y) &= h(x) - h(y) - \nabla h(y)(x - y), \\
&= \sum_{j=1}^n x_j \log(x_j) - \sum_{j=1}^n y_j \log(y_j) - \sum_{j=1}^n (\log(x_j) + 1)(x_j - y_j), \\
&= \sum_{j=1}^n x_j \log(x_j) - \sum_{j=1}^n y_j \log(y_j) - \sum_{j=1}^n x_j \log(y_j) + \sum_{j=1}^n y_j \log(y_j) \\
&\quad + \sum_{j=1}^n (x_j - y_j), \\
&= \sum_{j=1}^n x_j \log(x_j) - \sum_{j=1}^n x_j \log(y_j) + \sum_{j=1}^n (x_j - y_j), \\
&= \sum_{j=1}^n x_j (\log(x_j) - \log(y_j)) + \sum_{j=1}^n (x_j - y_j), \\
&= \sum_{j=1}^n [x_j \log\left(\frac{x_j}{y_j}\right) + (x_j - y_j)]. \text{ (ver figura 2.5 para } n=1)
\end{aligned}$$

Observemos que  $h(x)$  es frontera coerciva y zona coerciva.

$$\partial S = \{(x_1 \cdots x_n) : x_i = +\infty, 0 \text{ con } 1 \leq i \leq n\}$$

$$\begin{aligned}
\nabla h(y)(x - y) &= (\log(y_1) + 1, \dots, \log(y_n) + 1)(x - y), \\
&= \sum_{j=1}^n (\log y_j + 1)(x - y_j).
\end{aligned}$$

Si  $y_j = +\infty$  y  $\{y^k\}$  para algún  $1 \leq j \leq n$ , una sucesión tal que  $\lim_{k \rightarrow +\infty} y^k = y_j$  la expresión  $(\log y_j + 1)(x - y_j)$  cuando sustituimos  $y_j$  por los valores de la sucesión  $\{y^k\}$  obtenemos una sucesión con límite  $-\infty$ , en efecto.

$$\begin{aligned}
\lim_{k \rightarrow -\infty} (\log(y^k) + 1)(x - y^k) &= \lim_{k \rightarrow +\infty} (\log(y^k) + 1) \lim_{k \rightarrow +\infty} (x - y^k), \\
&= (\infty)(-\infty), \\
&= -\infty.
\end{aligned}$$

Si  $x_j = 0$  y  $\{y^k\}$  para algún  $1 \leq j \leq n$ , una sucesión tal que  $\lim_{k \rightarrow +\infty} y^k = x_j$  la expresión  $(\log y_j + 1)(x - y_j)$  cuando sustituimos  $y_i$  por los valores de la sucesión  $\{y^k\}$

obtenemos una sucesión con limite  $-\infty$  en efecto.

$$\begin{aligned} \lim_{k \rightarrow -\infty} (\log y_j + 1)(x - y_k) &= \lim_{k \rightarrow +\infty} (\log(x^k) + 1) + \lim_{k \rightarrow +\infty} (x - y^k) \\ &= (-\infty)(x) \\ &= -\infty. \end{aligned}$$

Ya con esto se tiene que si  $\bar{x} \in \partial S$  es por que una de sus componentes es 0 o  $+\infty$ . Con estos se generar una sucesión tal que converge a  $\bar{x}$  y por lo anterior  $\lim_{k \rightarrow +\infty} \nabla h(y^k)(x - y^k)$  es igual a  $-\infty$  con lo que  $h(x)$  es frontera coerciva.

Sea  $y \in \mathbb{R}^n$  y como  $\nabla h(x) = \log(x) + 1$ , igualando  $y$  con  $\nabla h(x)$  tenemos

$$\begin{aligned} y &= \nabla h(x), \\ (y_1, \dots, y_n) &= (\log(x_1) + 1, \dots, \log(x_n) + 1). \end{aligned}$$

Así que  $y_j = \log(x_j) + 1$  con lo que  $x_j = \exp(y_j + 1)$  si tomamos los  $x = (x_1, \dots, x_n)$  de esta manera tenemos que para cada  $y \in \mathbb{R}^n$  existe  $x$  tal que  $\nabla h(x) = y$  con lo que  $h(x)$  es zona coerciva.

Gráfica de  $h(x)$  y  $D_h(x, y)$  en  $S = \mathbb{R}_{++}$

Ahora presentaremos el método de punto proximal con distancias de Bregman.

El problema de interés es

$$\begin{aligned} \text{mín } f(x), \\ \text{s.a } x \in \bar{S}. \end{aligned} \tag{2.34}$$

Con  $S \subset \mathbb{R}^n$  un conjunto convexo y abierto,  $\bar{S}$  es la clausura de  $S$  y  $f$  convexa y continua en  $\bar{S}$ . El Método de punto proximal con distancia de Bregman esta definido como.

$$x^0 \in S, \tag{2.35}$$

$$x^{k+1} = \operatorname{argmin}\{f(x) + \lambda_k D_h(x, y) : x \in \bar{S}\}. \tag{2.36}$$

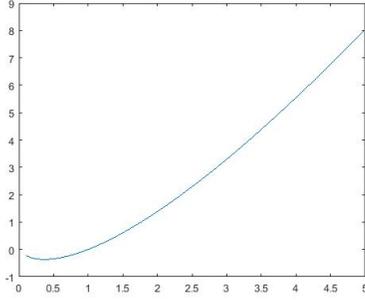


Figura 2.4:  $h(x) = x \log(x)$ .

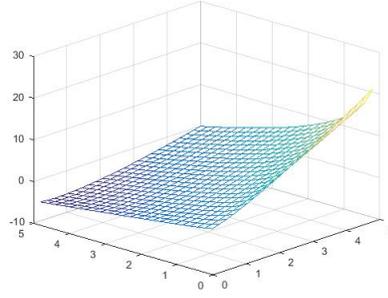


Figura 2.5:  $D_h(x, y) = x \log\left(\frac{x}{y}\right) + (x - y)$ .

Donde  $h$  es una función Bregman con zona  $S$  y  $\lambda_k$  satisfaciendo  $0 < \lambda_k < \bar{\lambda}$  para algún  $\bar{\lambda} > 0$  Asumiendo que el problema (2.63) tiene solución y que  $f$  es acotado en  $\bar{S}$  por debajo procedamos a el análisis de convergencia del método de punto proximal con distancia de Bregman.

**Teorema 2.3** Ver[6] [teorema 10.1]

Si el problema (2.34) tiene solución y  $h$  es límite coerciva con respecto a  $S$  entonces la sucesión  $\{x^k\}$  generada por (2.35) (2.36) converge a la solución  $x^*$  del problema (2.34)

**Demostración:**

Se seguirá la misma línea de la demostración del teorema 2.1, la cual consta de cuatro pasos

**Paso 1** verificar que la sucesión está bien definida y contenida en  $S$ .

Sea  $\beta$  una cota inferior de  $f$  en  $\bar{S}$  y  $f_k(x) = f(x) + \lambda_k D_h(x, x^k)$  con lo que  $f_k(x) \geq \beta + \lambda_k D_h(x, x^k)$  y por propiedad de  $(B_3)$  tenemos que  $f_k(x)$  es acotada por debajo así que la minimización en (2.36) se reduce a evaluar  $f_k(x)$  sobre conjuntos compactos, ya que  $f_k(x)$  es estrictamente convexa por la convexidad de  $h$  y la estrictamente convexidad de  $D_h(\cdot, y)$  demostrada en *iii*) de la proposición con lo que el mínimo es único.

Veamos que  $x^{k+1} \in S$

De (2.36) tenemos

$$\begin{aligned}
0 &= \partial f_k(x^{k+1}), \\
&= \partial(f(x^{k+1}) + \lambda_k D_h(x^{k+1}, x^k)), \\
&= \partial f(x^{k+1}) + \lambda_k \partial D_h(x^{k+1}, x^k), \\
&= \partial f(x^{k+1}) + \lambda_k (\partial h(x^{k+1}) - \partial h(x^k)^t).
\end{aligned}$$

Lo cual implica que

$$\lambda_k \partial h(x^k) = \partial f(x^{k+1}) + \lambda_k \partial h(x^{k+1})^t, \quad (2.37)$$

$$= \partial(f(x) + \lambda_k h(x))(x^{k+1}) \Rightarrow \lambda_k \partial h(x^k) \in \partial(f(x) + \lambda_k h(x))(x). \quad (2.38)$$

Bajo la hipótesis de que la sucesión es frontera coerciva,  $\partial(f + \lambda_k h)(x) = \emptyset$ , para todo  $x \in \partial S$ , lo cual implica por (2.38) que  $x^{k+1} \in S$

Se tiene  $x \in \partial S$  y suponga que existe  $\xi \in \partial(f + \lambda_k h)(x)$ , tomemos  $z \in S$  se puede definir:

$$y^l = (1 - \varepsilon_l)x + \varepsilon_l z. \quad (2.39)$$

Con  $\lim_{l \rightarrow \infty} \varepsilon_l = 0$ , entonces  $y^l \in S$ , por convexidad de  $S$  y  $\lim_{l \rightarrow \infty} y^l = x$ , también

$$\begin{aligned}
\varepsilon_l \xi^t(z - x) &= \varepsilon_l (y^l - x), \text{ por (2.39)} \\
&\leq f(y^l) - f(x) + \lambda_k (h(y^l) - h(x)), \\
&\leq f(y^l) - f(x) + \lambda_k \nabla h^t(y^l - x), \\
&\leq \varepsilon_l (f(y^l) - f(x)) + \lambda_k \frac{\varepsilon_l}{1 - \varepsilon_l} \nabla h^t(y^l - x).
\end{aligned} \quad (2.40)$$

Por (2.40)

$$\frac{1 - \varepsilon_l}{\lambda_k} [f(x) - f(z) + \xi^t(z - x)] \leq \nabla h(y^l)^t(z - y^l). \quad (2.41)$$

Ya que  $\lim_{l \rightarrow \infty} y^l = x \in \partial S$ , por ser la sucesión frontera coerciva implica que la parte derecha de inecuación (2.41) tiende a  $-\infty$  cuando  $l$  tiende a  $\infty$  mientras que la inecuación izquierda tiene un límite finito, esta contradicción implica que  $\partial(f + \lambda_k h) = \emptyset$ , para todo  $x \in \partial S$  y también  $x^{k+1} \in S$ .

**Paso 2**  $D_h(\bar{x}, x^{k+1}) \leq D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k)$  para todo  $k$  y cada solución de (2.34).

Por el ítem *i*) de la proposición 2.2 se tiene que

$$D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^{k+1} - \bar{x} \rangle.$$

Por (2.36)

$$0 \in \partial[f + \lambda_k D_h(\cdot, x^k)](x^{k+1}).$$

Si  $y^k = \lambda_k[\nabla h(x^k) - \nabla h(x^{k+1})]$  entonces

$$D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \frac{1}{\lambda_k} \langle y^k, x^{k+1} - x^k \rangle, \quad (2.42)$$

$$\geq \frac{1}{\lambda_k} (f(x^{k+1}) - f(\bar{x})) \quad (2.43)$$

$$\geq 0. \quad (2.44)$$

**Paso 3** Observemos que  $x^k$  es acotada y  $\lim_{k \rightarrow \infty} x^{jk} = \hat{x}$  implica que  $\lim_{k \rightarrow \infty} x^{jk+1} = \hat{x}$ .

Por el paso 2  $\{D_h(\bar{x}, x^k)\}$  es decreciente y no negativa, convergente y  $D_h(x^{k+1}, x^k) \leq D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1})$  se tiene que

$$\lim_{k \rightarrow \infty} D_h(x^{k+1}, x^k) = 0 \quad (2.45)$$

Ya que  $\{D_h(\bar{x}, x^k)\}$  es decreciente se tiene que  $D_h(\bar{x}, x^k) < D_h(\bar{x}, x^0)$  y por lo tanto  $\{x^k\}$  es acotada y por  $B_3$  si  $\lim_{k \rightarrow \infty} x^{jk} = \hat{x}$ , para toda sucesión  $\{x^{jk}\}$  de  $\{x^k\}$ , entonces  $\lim_{k \rightarrow \infty} x^{jk+1} = \hat{x}$  por  $B_5$ .

**Paso 4** La sucesión  $\{x^k\}$  tiene un punto de acumulación todo los cuales son solución de (2.34). Tomando una solución  $\bar{x}$  de (2.34) sea  $\hat{x}$  un punto de acumulación de

$\{x^k\}$  y  $\{x^{j^k}\}$  una subsucesión de  $\{x^k\}$  veamos que  $\lim_{k \rightarrow \infty} x^{j^k} = \hat{x}$ . la existencia de  $\hat{x}$  se tiene del poso 3 el cual también asegura que  $\lim_{k \rightarrow \infty} x^{j^{k+1}} = \hat{x}$ . De (2.81)

$$\begin{aligned} 0 &\leq \frac{1}{\lambda} (f(x^{k+1}) - f(\bar{x})), \\ &\leq \frac{1}{\lambda_k} (f(x^{j^{k+1}}) - f(\bar{x})), \\ &\leq D_h(\bar{x}, x^{j^k}) - D_h(\bar{x}, x^{j^{k+1}}) - D_h(x^{j^{k+1}}, x^k) \rightarrow 0 \text{ cuando } k \rightarrow \infty. \end{aligned}$$

Usando la convergencia de  $\{D_h(\bar{x}, x^k)\}$  y la frontera coersiva. Tomando con  $k$  tendiendo a  $\infty$   $f(\hat{x}) = f(\bar{x})$  ya que  $\bar{S}$  es cerrado y  $\{x^k\} \in \bar{S}$ , se tiene que  $\hat{x} \in \bar{S}$  y  $\hat{x}$  resuelve (2.34).

Ahora se necesita un teorema de Féjer convergencia para distancias de Bregman, el cual de hecho se tiene pero no podemos usarlo directamente.

Sea  $\hat{x}$  un punto de acumulación de  $\{x^k\}$  y tomando una subsucesión  $\{x^{j^k}\}$  de  $\{x^k\}$  tal que  $\lim_{k \rightarrow \infty} x^k = \hat{x}$ .

Por  $B_4$   $\lim_{k \rightarrow \infty} D_h(\hat{x}, x^{j^k}) = 0$  y por paso 4  $\hat{x}$  es solución de (2.34) y por paso 2  $\{D_h(\hat{x}, x^k)\}$  es no negativa y decreciente con lo cual es una subsucesión convergente y convergiendo a cero.

Por  $B_4$  se tiene que  $\lim_{k \rightarrow \infty} x^k = \hat{x}$ .

## 2.5. $\varphi$ -divergencia

En esta sección se discutirá otra clase de casi-distancia, la cual se llamara  $\varphi$ -divergencia denotada por  $d_\varphi(\cdot, \cdot)$  definida en el octante positivo de  $\mathbb{R}^n$  (es decir  $\mathbb{R}_{++}^n := x \in \mathbb{R}^n : x > 0$ ). Tomando  $\varphi : \mathbb{R}_{++} \rightarrow \mathbb{R}$  convexa y tres veces continuamente diferenciable, satisfaciendo

$$\varphi(1) = \varphi'(1) = 0, \quad \varphi''(1) > 0 \quad \lim_{t \rightarrow 0} \varphi'(t) = -\infty. \quad (2.46)$$

**Definición 2.4** (ver [6] Definición 12.1)

Si  $\varphi$  satisface (2.86) se define la función  $d_\varphi : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}$  llamada  $\varphi$ -divergencia y dada por:

$$d_\varphi(x, y) = \sum_{i=1}^n y_i \varphi \left( \frac{x_i}{y_i} \right). \quad (2.47)$$

Ahora presentaremos algunos ejemplos de funciones  $\varphi$  y sus correspondientes  $\varphi$ -divergencia.

**Ejemplo 2.3**  $\varphi_1(t) = t \log(t) - t + 1$ , (ver Figura 2.6)

Primero observemos que  $\varphi_1$  satisface la condiciones (2.46).  $\varphi_1$  es tres veces diferenciable ya que es suma de funciones que son  $C^\infty$ .

Calculando  $\nabla^2 \varphi_1(t)$

$\nabla \varphi_1(t) = \log(t)$ ;  $\nabla^2 \varphi_1(t) = \frac{1}{t}$  como  $t \in \mathbb{R}_{++}$   $\frac{1}{t} > 0$  y por la condición b) de la proposición 1.1  $\varphi_1(t)$  es estrictamente convexa.

$$\varphi_1(t) = t \log(t) - t + 1 \Rightarrow \varphi_1(1) = 1 \log(1) - 1 + 1 = \log(1) = 0,$$

$$\varphi_1'(t) = \log t \Rightarrow \varphi_1'(1) = \log(1) = 0,$$

$$\varphi_1''(t) = \frac{1}{t} \Rightarrow \varphi_1''(1) = \frac{1}{1} > 0,$$

$$\lim_{t \rightarrow 0} \log(t) = -\infty.$$

$$\begin{aligned} d_{\varphi_1}(x, y) &= \sum_{i=1}^n y_i \varphi_1 \left( \frac{x_i}{y_i} \right), \\ &= \sum_{i=1}^n y_i \left( \frac{x_i}{y_i} \log \left( \frac{x_i}{y_i} \right) - \frac{x_i}{y_i} + 1 \right), \\ &= \sum_{i=1}^n x_i \log \left( \frac{x_i}{y_i} \right) + y_i - x_i. \quad (\text{ver Figura 2.7 para } n=1) \end{aligned}$$

Esta función  $d_{\varphi_1}$  es denominada la  $\varphi$ -divergencia de Kullback-Leibler y puede por lo tanto ser extendida a  $\mathbb{R}_+^n \times \mathbb{R}_{++}^n$ .

Si  $h_1 = \sum_{i=1}^n x_i \log(x_i)$  se tiene que  $D_{h_1} = d_{\varphi_1}$ .

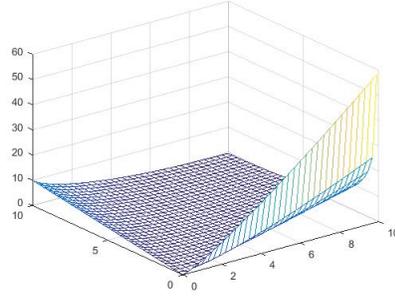
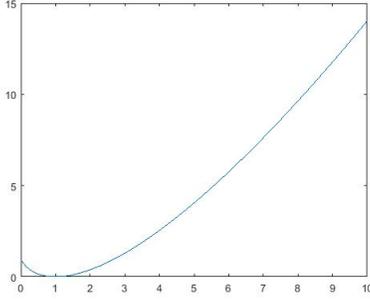


Figura 2.6:  $\varphi_2(t) = t \log(t) - t + 1$ .

Figura 2.7:  $d_{\varphi_2}(x, y) = x \log\left(\frac{x}{y}\right) + y - x$ .

**Ejemplo 2.4**  $\varphi_2(t) = t - \log(t) - 1$ , (ver Figura 2.8)

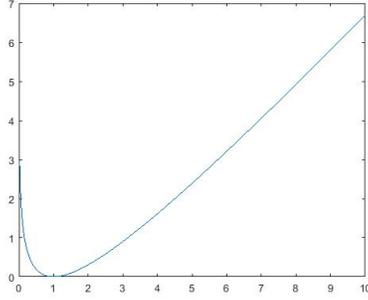
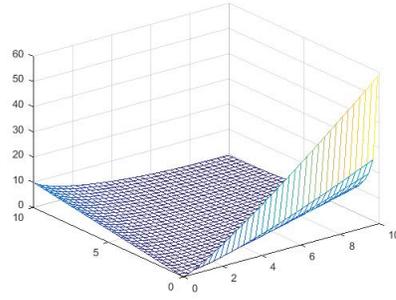
Al igual que en el ejemplo anterior, observe que se satisfacen las condiciones (2.46)  $\varphi_2(t)$  es tres veces diferenciable, ya que, es suma de funciones que son  $C^\infty$ .

La convexidad de  $\varphi_2(t)$ , se tiene por que es suma de funciones convexas.  $\varphi_2(t) = t - \log(t) - 1 \Rightarrow \varphi_2(1) = 1 - \log(1) - 1 = \log(1) = 0$ ,  
 $\varphi_2'(t) = \frac{t-1}{t} \Rightarrow \varphi_2'(1) = \frac{1-1}{1} = 0$ ,  
 $\varphi_2''(t) = \frac{1}{t^2} \Rightarrow \varphi_2''(1) = \frac{1}{1} > 0$ ,  
 $\lim_{t \rightarrow 0} \frac{t-1}{t} = -\infty$ .

$$\begin{aligned} d_{\varphi_2}(x, y) &= \sum_{i=1}^n y_i \left( \frac{x_i}{y_i} - \log\left(\frac{x_i}{y_i}\right) - 1 \right), \\ &= \sum_{i=1}^n x_i \log\left(\frac{x_i}{y_i}\right) + y_i - x_i, \text{ (ver Figura 2.9 para } n=1\text{)} \\ &= d_{\varphi_1}(y, x). \end{aligned}$$

**Ejemplo 2.5**  $\varphi_3(t) = (\sqrt{t} + 1)^2$ , (ver Figura 2.10)

Como en el ejemplo anterior observe que se satisfacen las condiciones (2.86)  $\varphi_3$  es tres veces diferenciable, ya que, es compuesta de funciones de clase  $C^\infty$

Figura 2.8:  $\varphi_2(t) = t - \log(t) - 1$ .Figura 2.9:  $d_{\varphi_2}(x, y) = x \log\left(\frac{x}{y}\right) + y - x$ .

Calculando  $\nabla^2 \varphi_3(t)$ .

$\nabla \varphi_3(t) = 4 - \frac{4}{\sqrt{t}}$ ;  $\nabla^2 \varphi_3(t) = \frac{1}{t^{3/2}}$  como  $t \in \mathbb{R}_{++}$ ,  $\frac{1}{t^{3/2}} > 0$  y por la parte b) de la proposición 1.1,  $\varphi_3$  es una función estrictamente convexa

$$\varphi_3(t) = (\sqrt{t} - 1)^2 \Rightarrow \varphi_3(1) = (\sqrt{1} - 1)^2 = 0,$$

$$\varphi_3'(t) = 4 - \frac{4}{\sqrt{t}} \Rightarrow \varphi_3'(1) = 4 - \frac{4}{\sqrt{1}} = 0,$$

$$\varphi_3''(t) = \frac{1}{t^{3/2}} \Rightarrow \varphi_3''(1) = 4 - \frac{1}{1^{3/2}},$$

$$\lim_{t \rightarrow 0} \frac{4\sqrt{t} - 4}{\sqrt{t}} = -\infty.$$

$$\begin{aligned} d_{\varphi_3}(x, y) &= \sum_{i=1}^n y_i \left( \sqrt{\frac{x_i}{y_i}} - 1 \right)^2, \\ &= \sum_{i=1}^n y_i \left( \frac{x_i}{y_i} - 2\sqrt{\frac{x_i}{y_i}} + 1 \right), \\ &= \sum_{i=1}^n y_i (y_i - 2\sqrt{x_i y_i} + y_i), \\ &= \sum_{i=1}^n (\sqrt{x_i} - \sqrt{y_i})^2. \quad (\text{ver Figura 2.11 para } n=1) \end{aligned}$$

La siguiente proposición establece unas propiedades que satisfacen la función  $\varphi$ -divergencia.

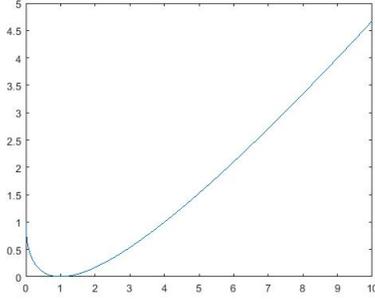


Figura 2.10:  $\varphi_3(t) = (\sqrt{t} + 1)^2$

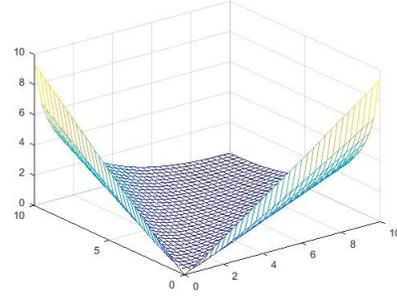


Figura 2.11:  $d_{\varphi_3}(x, y) = (\sqrt{x} - \sqrt{y})^2$

**Proposición 2.3** (ver [6] Proposición 12.1)

Sea  $\varphi$  satisfaciendo las condiciones (2.46)

- i)  $d_{\varphi}(x, y) \geq 0$  para todo  $x, y \in \mathbb{R}_{++}^n$ ,
- ii)  $d_{\varphi}(x, y) = 0$  sii  $x = y$ ,
- iii) El conjunto de nivel de  $d_{\varphi}(\cdot, y)$  es acotada para toda  $y \in \mathbb{R}_{++}^n$ ,
- iv) El conjunto de nivel de  $d_{\varphi}(x, \cdot)$  es acotado para todo  $x \in \mathbb{R}_{++}^n$ ,
- v)  $d_{\varphi}(x, y)$  es continuamente convexa en  $x, y$  y estrictamente convexa en  $x$ ,
- vi)  $\lim_{k \rightarrow \infty} d_{\varphi}(y, y^k) = 0$  sii  $\lim_{k \rightarrow \infty} y^k = y$ ,

**Demostración:**

Veamos que bajo las condiciones (2.46) de  $\varphi$  esta es no negativa. como  $\varphi'(1) = 0$  y  $\varphi''(1) > 0$  por el teorema anterior 0 es un mínimo local estricto, como 0 ya es un mínimo local estricto, por el teorema anterior 0 es un mínimo global a si que  $0 \leq \varphi(x) \forall x \in \mathbb{R}_{++}$ .

i)

Como  $\varphi$  es no negativa tenemos que

$$\begin{aligned} \varphi\left(\frac{x_i}{y_i}\right) &\geq 0 \text{ y como } y_i > 0 \Rightarrow y_i \varphi\left(\frac{x_i}{y_i}\right) \geq 0. \\ &\Rightarrow \sum_{i=1}^n y_i \varphi\left(\frac{x_i}{y_i}\right) \geq 0, \\ &\Rightarrow d_{\varphi}(x, y) \geq 0. \end{aligned}$$

**ii)** $(\Rightarrow)$ 

$$\begin{aligned}d_{\varphi}(x, y) = 0 &\Rightarrow \sum_{i=1}^n y_i \varphi\left(\frac{x_i}{y_i}\right) = 0, \\&\Rightarrow y_i \varphi\left(\frac{x_i}{y_i}\right) = 0, \\&\Rightarrow \varphi\left(\frac{x_i}{y_i}\right) = \frac{0}{y_i}, \\&\Rightarrow \varphi\left(\frac{x_i}{y_i}\right) = 0, \\&\Rightarrow \frac{x_i}{y_i} = 1 \quad (\text{por ser } 0 \text{ el m\u00ednimo}), \\&\Rightarrow x_i = y_i \quad (\forall i = 1, \dots, n), \\&\Rightarrow x = y.\end{aligned}$$

 $(\Leftarrow)$ 

$$\begin{aligned}x = y &\Rightarrow x_i = y_i \quad \forall i = 1, \dots, n, \\&\Rightarrow \frac{x_i}{y_i} = 1 \quad \text{ya que } y_i \neq 0 \quad \forall i = 1, \dots, n, \\&\Rightarrow \varphi\left(\frac{x_i}{y_i}\right) = 0, \\&\Rightarrow y_i \varphi\left(\frac{x_i}{y_i}\right) = 0, \\&\Rightarrow \sum_{i=1}^n y_i \varphi\left(\frac{x_i}{y_i}\right) = 0, \\&\Rightarrow d_{\varphi}(x, y) = 0.\end{aligned}$$

**iii)**

Tenemos que  $d_\varphi(\cdot, y) = \{e \in \mathbb{R}_{++}^n : d_\varphi(x, y) \leq \alpha\}$  para todo  $y \in \mathbb{R}_{++}^n$

$$\begin{aligned}
d_\varphi(x, y) \leq \alpha &\Rightarrow \sum_{i=1}^n y_i \varphi\left(\frac{x_i}{y_i}\right) \leq \alpha, \\
&\Rightarrow \|y\| \left\| \left( \varphi\left(\frac{x_1}{y_1}\right), \dots, \varphi\left(\frac{x_1}{y_2}\right) \right) \right\| \cos(\theta) \leq \alpha, \\
&\Rightarrow \|y\| \left\| \left( \varphi\left(\frac{x_1}{y_1}\right), \dots, \varphi\left(\frac{x_1}{y_2}\right) \right) \right\| < \alpha, \\
&\Rightarrow 0 \leq \|y\| \left\| \left( \varphi\left(\frac{x_1}{y_1}\right), \dots, \varphi\left(\frac{x_1}{y_2}\right) \right) \right\| < \alpha \text{ y que } d_\varphi(x, y) \geq 0, \\
&\Rightarrow 0 \leq \left\| \left( \varphi\left(\frac{x_1}{y_1}\right), \dots, \varphi\left(\frac{x_1}{y_2}\right) \right) \right\| < \frac{\alpha}{\|y\|} \text{ ya que } y \neq 0, \\
&\Rightarrow 0 \leq \varphi\left(\frac{x_i}{y_i}\right) < \frac{\alpha}{\|y\|} \text{ siendo } \|\cdot\| \text{ la norma del maximo.} \\
&\Rightarrow 0 \leq \varphi\left(\frac{x_i}{y_i}\right) < \frac{\alpha}{\|y\|} \text{ por ser } \varphi \text{ no negativa.}
\end{aligned}$$

Como el conjunto  $\left[0, \frac{\alpha}{\|y\|}\right]$  es cerrado y de topología se sabe que la imagen inversa de cerrados mediante una función continua es cerrado como  $\varphi$  lo es la imagen inversa de  $\left[0, \frac{\alpha}{\|y\|}\right]$  mediante  $\varphi$  es cerrado, además tenemos que es compacto  $x_i$  es acotado para todo  $i = 1, \dots, n$  en consecuencia  $x$  es acotado.

v)

Como  $y_i$  es continua y  $\varphi(x)$  son continuas, se tiene que  $y_i \varphi\left(\frac{x_i}{y_i}\right)$  es continua y como la suma de funciones continuas es continua tenemos que  $\sum_{i=1}^n y_i \varphi\left(\frac{x_i}{y_i}\right)$  es continua así que  $d_\varphi(x, y)$  es continua.

Sea  $x$  fijo y  $a, b \in \mathbb{R}_{++}$

$$d_\varphi(x, \alpha a + (1 - \alpha)b) = \sum_{i=1}^n (\alpha a_i + (1 - \alpha)b_i) \varphi\left(\frac{x_i}{\alpha a_i + (1 - \alpha)b_i}\right).$$

Sea  $y$  fijo y  $a, b \in \mathbb{R}_{++}$ ,

$$\begin{aligned}
 d_\varphi(\alpha a + (1 - \alpha)b, y) &= \sum_{i=1}^n y_i \varphi\left(\frac{\alpha a_i + (1 - \alpha)b_i}{y_i}\right), \\
 &\leq \sum_{i=1}^n y_i \left( \alpha \varphi\left(\frac{a_i}{y_i}\right) + (1 - \alpha) \varphi\left(\frac{b_i}{y_i}\right) \right), \\
 &= \sum_{i=1}^n y_i \alpha \varphi\left(\frac{a_i}{y_i}\right) + \sum_{i=1}^n y_i (1 - \alpha) \varphi\left(\frac{b_i}{y_i}\right), \\
 &= \alpha \sum_{i=1}^n y_i \varphi\left(\frac{a_i}{y_i}\right) + (1 - \alpha) \sum_{i=1}^n y_i \varphi\left(\frac{b_i}{y_i}\right), \\
 &= \alpha d_\varphi(a, y) + (1 - \alpha) d_\varphi(b, y).
 \end{aligned}$$

Con lo que  $d_\varphi(x, y)$  es convexa en  $y$ .

vi)

( $\Rightarrow$ )

Sea  $y \in \mathbb{R}_{++}^n$  como  $\lim_{k \rightarrow \infty} d_\varphi(y, y^k) = 0$  se tiene que dado  $\varepsilon > 0$  existe  $k > 0, k \in \mathbb{N}$  tal que si  $m > k$

$$\begin{aligned}
 0 \leq |d_\varphi(y, y^m)| < \varepsilon &\Rightarrow 0 < \sum_{i=1}^n y_i \varphi\left(\frac{y_i^m}{y_i}\right) < \varepsilon, \\
 &\Rightarrow 0 < y_i \varphi\left(\frac{y_i^m}{y_i}\right) < \varepsilon \text{ para todo } i = 1, \dots, n, \\
 &\Rightarrow 0 < \varphi\left(\frac{y_i^m}{y_i}\right) < \frac{\varepsilon}{y_i} \text{ ya que } y_i \neq 0, \\
 &\Rightarrow \varphi\left(\frac{y_i^m}{y_i}\right) \rightarrow 0 \text{ con } \varepsilon \rightarrow 0, \\
 &\Rightarrow \frac{y_i^m}{y_i} \rightarrow 1 \text{ por ser 1 el \u00fanico m\u00ednimo en } \varphi, \\
 &\Rightarrow y_i^m \rightarrow y_i.
 \end{aligned}$$

( $\Leftarrow$ )

Sea  $y \in \mathbb{R}_{++}^n$  como  $\lim_{k \rightarrow \infty} y^k = y$  se tiene que dado  $\varepsilon > 0$  existe  $k > 0, k \in \mathbb{N}$  tal que si  $m > k$

$$\begin{aligned}
0 < |y_i - y_i^m| < \varepsilon &\Rightarrow 0 < \left| \frac{y_i^m - y_i}{y_i} \right| < \frac{\varepsilon}{y_i} \text{ para toda } i = 1, \dots, n, \\
&\Rightarrow 0 < \left| \frac{y_i^m}{y_i} - 1 \right| < \frac{\varepsilon}{y_i}, \\
&\Rightarrow \frac{y_i^m}{y_i} \rightarrow 0 \text{ cuando } \varepsilon \rightarrow 0, \\
&\Rightarrow \varphi\left(\frac{y_i^m}{y_i}\right) \rightarrow 0 \text{ ya que es continua,} \\
&\Rightarrow 0 < |y_i \varphi\left(\frac{y_i^m}{y_i}\right)| < y_i \varepsilon \text{ por convergencia,} \\
&\Rightarrow 0 < d_\varphi(y, y^m) < y_i \varepsilon, \\
&\Rightarrow d_\varphi(y, y^m) \rightarrow 0 \text{ ya que } y_i \text{ es constante con } \varepsilon \rightarrow 0.
\end{aligned}$$

Ahora presentaremos el algoritmo de punto proximal con  $\varphi$ -divergencia el problema de interés es

$$\begin{aligned}
&\text{mín } f(x) \\
&\text{s.a } x \geq 0
\end{aligned} \tag{2.48}$$

con  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  convexa, el algoritmo de punto proximal con  $\varphi$ -divergencia del problema (2.48) genera una  $\{x^k\} \subset \mathbb{R}^n$  dada por

$$x^0 > 0, \tag{2.49}$$

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) + \lambda_k d_\varphi(x, x^k)\}. \tag{2.50}$$

El siguiente ejemplo que se presentara muestra que el algoritmo de punto proximal con  $\varphi$ -divergencia no siempre genera una sucesión Féjer convergente.

**Ejemplo 2.6** Sea  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por  $f(x) = (x_1 - x_2)^2$  donde  $x_1$  y  $x_2$  son las componentes del vector  $x$  y  $\varphi(t)$  dada por

$$\varphi(t) = \begin{cases} \frac{1}{16} \left( \frac{1}{16} t^{-3} - \frac{17}{16} t^{-2} + 17t^{-1} - 6 \right) & \text{si } t \in (0, 1/4) \\ (t-1)^2 & \text{si } t \in [1/4, +\infty) \end{cases}$$

La definición de  $\varphi(t)$  en el conjunto  $(0, 1/4)$  es para asegurar que  $\varphi(t)$   $\varphi$ -divergencia y además, ya que  $\varphi(t)$  es tres veces continuamente diferenciable en  $t = 1/4$  y  $U = \{x \in \mathbb{R}^2 : x_1 = x_2\}$  es el conjunto solución del problema, Si tomamos  $\lambda_k = 1$  para toda  $k \in \mathbb{N}$  y  $x^0 = (1, 2)$  tenemos que

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}_{x \in \mathbb{R}^2} \{f(x) + d_\varphi(x, x^0)\}, \\ &= \operatorname{argmin}_{x \in \mathbb{R}^2} \left\{ (x_1 - x_2)^2 + 1\varphi\left(\frac{x_1}{1}\right) + 2\varphi\left(\frac{x_2}{2}\right) \right\}, \\ &= \operatorname{argmin}_{x \in \mathbb{R}^2} \left\{ (x_1 - x_2)^2 + 1(x_1 - 1)^2 + 2\left(\frac{x_2}{2} - 2\right)^2 \right\}. \end{aligned}$$

Como la  $(x_1 - x_2)^2 + 1(x_1 - 1)^2 + 2\left(\frac{x_2}{2} - 2\right)^2$  es diferenciable y por la condiciones de optimalidad para funciones diferenciable tenemos que

$$\nabla \left( (x_1 - x_2)^2 + 1(x_1 - 1)^2 + 2\left(\frac{x_2}{2} - 2\right)^2 \right) = 0. \quad (2.51)$$

Lo que implica el siguiente sistema

$$\begin{aligned} \begin{cases} 2(x_1 - x_2 + 2(x_1 - 1)) &= 0 \\ -2(x_1 - x_2 + 2\left(\frac{x_2}{2} - 1\right)) &= 0 \end{cases} &\Rightarrow \begin{cases} 4x_1 - 2x_2 &= 2 \\ -2x_1 + 3x_2 &= 2 \end{cases} \\ &\Rightarrow \begin{cases} x_1 &= 5/4 \\ x_2 &= 6/4 \end{cases} \end{aligned}$$

Con esto tenemos que  $x^1 = (5/4, 6/4)$  veamos que los elementos de la sucesión no son Féjer convergentes para  $u = (3, 3)$ , fácil de ver que  $u \in U$

$$\begin{aligned} d_\varphi(x^1, u) &= 3\left(\frac{5}{12} - 1\right)^2 + 3\left(\frac{1}{2} - 1\right)^2, \\ &= 3\left(\left(\frac{-7}{12}\right)^2 + \left(\frac{-1}{2}\right)^2\right), \\ &= \frac{49}{48} + \frac{3}{4}, \\ &= \frac{85}{48}. \end{aligned}$$

$$\begin{aligned}
d_\varphi(x^0, u) &= 3 \left( \frac{1}{3} - 1 \right)^2 + 3 \left( \frac{2}{3} - 1 \right)^2, \\
&= 3 \left( \left( \frac{-2}{3} \right)^2 + \left( \frac{1}{3} \right)^2 \right), \\
&= \frac{4}{3} + \frac{1}{3}, \\
&= \frac{5}{3}.
\end{aligned}$$

Con esto se tiene que  $d_\varphi(x^0, u) < d_\varphi(x^1, u)$  lo que implica que la sucesión generada no es Féjer convergente.

**Definición 2.5** (ver [6] Definición 13.1)

Una sucesión  $\{y^k\}$  es **casi-Féjer convergente** para un conjunto  $U \subset \mathbb{R}_{++}^n$  con respecto a la  $\varphi$ -divergencia  $d_\varphi$ . Si para cada  $u \in U$  existe una sucesión  $\{\varepsilon_k\} \subset \mathbb{R}_{++}$  tal que  $\sum_{k=0}^{\infty} \varepsilon_k < \infty$  y para todo  $k \geq 0$

$$d_\varphi(u, y^{k+1}) \leq d_\varphi(u, y^k) + \varepsilon_k.$$

**Proposición 2.4** (ver [6] Proposición 13.1)

Si  $\{y^k\} \subset \mathbb{R}_{++}^n$  es casi-Féjer convergente para  $U \subset \mathbb{R}_{++}^n$  con respecto a una  $\varphi$ -divergencia  $d_\varphi$  entonces  $\{y^k\}$  es acotada. Si un punto de acumulación de  $\bar{y}$  de  $\{y^k\}$  pertenece a  $U$  entonces  $\bar{y} = \lim_{k \rightarrow \infty} y^k$ .

**Demostración:**

Sea  $u \in U$  como  $\{y^k\} \subset \mathbb{R}_{++}^n$  es casi-Féjer convergente existe  $\{\varepsilon_i\} \subset \mathbb{R}_{++}^n$  tal que  $\sum_{k=0}^{\infty} \varepsilon_k < \infty$  y para todo  $k \geq 0$  se tiene que

$$\begin{aligned}
d_\varphi(u, y^k) &\leq d_\varphi(u, y^{k-1}) + \varepsilon_{k-1}, \\
&\leq d_\varphi(u, y^1) + \sum_{i=1}^{k-1} \varepsilon_{k-1}, \\
&\leq d_\varphi(u, y^1) + \sum_{i=1}^{\infty} \varepsilon_{k-1}, \\
&\leq d_\varphi(u, y^1) + \beta \text{ si } \beta = \sum_{i=1}^{\infty} \varepsilon_{k-1} \text{ y por proposición 1.3}
\end{aligned}$$

Sea  $\bar{y} \in U$  un punto de acumulación de  $\{y^k\}$ , por definición existe una subsección  $\{y^{j_k}\}$  de  $\{y^k\}$  tal que  $\lim_{j \rightarrow \infty} y^{j_k} = \bar{y} \iff \forall \varepsilon > 0 \exists \bar{k} \in \mathbb{N}$  tal que

$$n > \bar{k} \Rightarrow d_\varphi(\bar{y}, y^{nk}) < \frac{\varepsilon}{2},$$

Además como  $\{\varepsilon_i\}$  es convergente en  $\mathbb{R} \iff \forall \varepsilon > 0 \exists \hat{k}$  tal que,

$$m > \hat{k} \Rightarrow \sum_{i=\hat{k}}^{\infty} \varepsilon_i < \frac{\varepsilon}{2},$$

tomando  $k^* = \max\{\bar{k}, \hat{k}\}$  se tiene que

$$\begin{aligned} d_\varphi(\bar{y}, y^{(k^*+1)k}) &\leq d_\varphi(\bar{y}, y^{(k^*)k}) + \varepsilon_{k^*,k}, \\ &\leq \frac{\varepsilon}{2} + \sum_{i=\hat{k}}^{\infty} \varepsilon_i, \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}, \\ &\leq \varepsilon. \end{aligned}$$

Tomando  $\varepsilon \rightarrow 0$  se tiene que  $d_\varphi(\bar{y}, y^{(k^*+1)k}) = 0$  y si  $\rho = (k^* + 1)k$  se obtiene que  $\lim_{\rho \rightarrow \infty} y^\rho = \bar{y}$ . Además del ejemplo 2.6 donde la sucesión generada por el algoritmo de punto proximal no es Féjer convergente puede darse el caso en que la distancia  $d_\varphi$  la solución de (2.48)  $u$  pertenece a  $\partial\mathbb{R}_+^n$  y la distancia  $d_\varphi(u, x^k)$  no esta bien definida. sin embargo sobre algunas condiciones impuestas a  $\varphi$ ,  $\{x^k\}$  es casi-Féjer convergente para el conjunto solución de (2.48) con respecto a la Kullback-Leibler, la cual admite puntos en  $\partial\mathbb{R}_+^n$  para un primer argumento y para el cual en la proposición (2.4) se cumple que  $u \in \mathbb{R}_+^n$ , con el objetivo de resolver este problema considera la  $\{x^k\}$  generada por (2.49)-(2.50) con  $d_\varphi$  tal que  $u \in U$  se define  $\delta_k$  como

$$\delta_k = d_\varphi(u, x^{k+1}) - d_\varphi(u, x^k). \quad (2.52)$$

---

**Proposición 2.5** (ver [6] Proposición 13.3)

- i) Si  $\varphi'(t) \leq \varphi''(1) \log(t)$  y existe un  $\hat{\lambda}$  tal que  $\lambda_k \geq \hat{\lambda} > 0$  para todo  $k$  entonces  $\sum_{k=0}^{\infty} \delta_k < \infty$  (y también  $\{x^k\}$  es casi-Féjer convergente para  $U$ )
- ii) Si  $\frac{\varphi''(1)}{t} \leq \varphi'(t) \leq \varphi''(1) \log(t)$  entonces  $\delta_k \leq 0$  (y también  $\{x^k\}$  Féjer convergente y casi-Féjer convergente para  $U$ ).

La demostración de esta proposición es muy extensa con lo que se puede observar en [8].



---

### Método de punto proximal con casi-distancia generalizada

---

En esta capítulo se considerara una casi-distancias propuestas por **Campos Castillo Y Hernández**, [5], que contiene como degenerado la  $\varphi$ -divergencia. esta casi-distancia es definida utilizando el conjugado shift de funciones penalidad que no necesariamente tiene que pasar por el origen con pendiente 1 y sus funciones conjugadas permiten valores negativos. Para un caso particular **Campos Castillo Y Hernández**, [5], podríamos decir que logran crear una generalización de la casi-distancia de Kullback-Leibler, dicha generalización puede ser vista como la diferencia entre una función casi de Bregman y sus aproximaciones lineales. Para valores específicos de los argumentos, los resultados de convergencia del método de punto proximal aplicado en el problema dual con esta casi-distancia y el lagrangeano aumentado con la con shift de función penalidad son mostrados también en este capítulo.

Consideremos el problema de programación convexa definido por

$$f^* = \inf\{f_0(x) : f_i(x) \leq 0 \ i = 1, \dots, m\}. \quad (3.1)$$

Donde  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  para  $i = 1, \dots, m$  son funciones convexas propias y cerradas. El problema dual convexo asociado a el problema (3.1) es

$$d^* = \inf\{-d(\mu) : \mu \geq 0\}. \quad (3.2)$$

Donde  $d(\mu) = \inf\{L(x, \mu) : x \in \mathbb{R}^n\}$  y

$$L(x, \mu) = f_0(x) + \sum_{i=1}^m \mu_i f_i(x). \quad (3.3)$$

Es la función lagrangeana.

Suponga que se tienen las siguientes hipótesis.

$H_1$ ) El conjunto optimal del problema (3.1) es no vacío y compacto.

$H_2$ ) Existe  $\bar{x}$  tal que  $f_i(\bar{x}) < 0$  para  $i = 1, \dots, m$ .

El problema (3.1) puede ser resuelto usando diferentes enfoques, por ejemplo se puede usar el método de lagrangeano aumentado, el cual puede ser introducido desde el punto de vista primal o desde un punto de vista dual, donde el método de multiplicador es construido aplicando la teoría de dualidad de Fenchel para el método de punto proximal.

El método de punto proximal con  $\varphi$ -divergencia puede resolver (3.2) generando una sucesión  $\{\mu^k\} \in \mathbb{R}_{++}^n$  tal que  $\mu^0 \in \mathbb{R}_{++}^n$  y

$$\mu^{k+1} = \inf\{-d(\mu) + r^k d_\varphi(\mu, \mu^k)\}. \quad (3.4)$$

Donde  $0 < \underline{r} < r^k < \bar{r} \leq 1$ ,

$$d_\varphi : \mathbb{R}_+^m \times \mathbb{R}_{++}^m \rightarrow \mathbb{R} \text{ es dado por } d_\varphi(s, \mu) = \sum_{i=1}^m \mu_i \varphi\left(\frac{s_i}{\mu_i}\right). \quad (3.5)$$

Tal que  $\varphi$  tenga las propiedades de la  $\varphi$  estudiada. Otras casi-distancias tienen su uso en diferentes contextos con diferentes propiedades y resultados, como la casi-distancia estudiada en [2], que es homogéneas de segundo orden y es dada por

$$d_\varphi(s, \mu) = \sum_{i=1}^m \mu_i^2 \varphi\left(\frac{s_i}{\mu_i}\right). \quad (3.6)$$

Por lo estudiado en la sección de distancia de Bregman el problema (3.2) puede ser resuelto usando iteraciones en (2.3) y cambiando  $d_\varphi$  por  $D_h$  **Campos, Castillo y Hernández**, [5] presentan una casi distancia para resolver el problema (3.2) la cual tiene propiedades interesantes y es dada por

$$d_{\varphi^*}^p(s, \mu) = \sum_{i=1}^m \left[ \frac{\mu_i^p}{c} \theta^*\left(\frac{cs_i}{\mu_i}\right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c) (s_i - \mu_i) \right]. \quad (3.7)$$

Donde  $\theta^*$  es la función conjugada de la función de penalidad  $\theta$  la cual no necesariamente pasa por el origen y satisface  $\theta'(0) = k$ ,  $p \geq 0$  y  $c \in \mathbb{R}_{++}$  tal que  $(\theta^*)'(c) = y \in \mathbb{R}$ . Para  $\mu \in \mathbb{R}_{++}^n$  fijo, si se define  $h_\mu : \mathbb{R}_+^m \rightarrow \mathbb{R}$  por  $h_\mu(s) = \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^*\left(\frac{cs_i}{\mu_i}\right)$  se tiene  $d_{\theta^*}^p(s, \mu) = h_\mu(s) - h_\mu(\mu) - \nabla h_\mu(\mu)^t (s - \mu)$  es decir,  $d_{\theta^*}^p$  puede ser vista como la diferencia entre  $h_\mu(s)$  y su aproximación en  $s = \mu$ . Observe que  $\theta^*$  es una traslación de la función  $\varphi$  usada en la  $\varphi$ -divergencia, en este caso con punto minimal en  $(k, \theta^*(k))$  y  $\theta^*(k) \in \mathbb{R}$ .

### 3.1. Generando la casi distancia

Una de las particularidades que tiene esta casi-distancia presentada es que esta utiliza una familia de penalidad que no solamente se restringe a pasar por el origen con pendiente uno, esto viene a ser una novedad en familia de penalización, ya que no había sido estudiada este tipo de penalidades hasta el momento en que se presentó dicha casi-distancia, además se mostrará que la casi-distancia presentada no es más que la conjugada de la función shift de función penalidad, dicha shift de función penalidad es más que una especie de traslación de las funciones de penalidad estudiadas aquí.

### 3.1.1. Funciones de penalidad

Las funciones de penalidad estudias es una familia de funciones sobre las que se definió la casi-distancia presentada por **Campos, Castillo y Hernández**, [5], que poseen características que nos permiten la convergencia del método de punto proximal con esta casi-distancia presentada.

Sea  $\theta$  una función estrictamente convexa dos veces diferenciable y estrictamente convexa definida en  $(-\infty, b)$  con  $0 < b \leq +\infty$  tal que satisfaga las siguientes condiciones:

- 1a)  $0 < \theta'(0) = k < +\infty$ .
- 2a)  $\lim_{t \rightarrow b} \theta'(t) = +\infty$ .
- 3a)  $\lim_{t \rightarrow -\infty} \theta'(t) = 0$ .
- 4a)  $\theta'' \geq \frac{1}{M}, \forall t \geq k = \theta'(0)$ .

Asociado con la función  $\theta$  nosotros consideraremos y denotaremos sus funciones conjugadas las cuales satisfacen las siguientes propiedades:

- 1b)  $\theta^*$  es una función diferenciable estrictamente convexa en  $(0, +\infty)$ .
- 2b)  $\theta^*$  es decreciente en  $(0, k)$  y creciente en  $(k, +\infty)$  con  $\theta^*(k) \in \mathbb{R}$ .
- 3b)  $\lim_{t \rightarrow 0^+} (\theta^*)'(t) = -\infty$  y  $\lim_{t \rightarrow +\infty} (\theta^*)'(t) = +\infty$ .
- 4b)  $\theta'' \geq M, \forall t \in [0, b]$  y para algún  $M > 0$ .

A continuación se presenta un ejemplo de una función  $\theta$  satisfaciendo las condiciones 1a)-4a), además de su función conjugada y sus respectivas gráficas.

**Ejemplo 3.1** Para  $\theta(t) = e^{1+t}$ . (ver Figura 3.1)

Se tiene que:

$$\theta(0) = e^{0+1} = e^1 = e,$$

$$\theta'(t) = e^{t+1},$$

$$\theta'(0) = e^{0+1} = e.$$

Como:

$$\theta^*(t) = \sup_{x \in \mathbb{R}} \{\langle t, x \rangle - \theta(x)\} = \sup_{x \in \mathbb{R}} \{tx - e^{x+1}\}.$$

Y teniendo que  $tx - e^{x+1}$  es diferenciable respecto a  $x$  e igualando la derivada a cero para calcular sus punto crítico

$$\begin{aligned} t - e^{t+1} = 0 &\Rightarrow t = e^{x+1}, \\ &\Rightarrow \ln(t) = x + 1, \\ &\Rightarrow x = \ln(t) - 1. \end{aligned}$$

Sustituyendo  $x$  en  $tx - e^{x+1}$  tenemos

$$t(\ln(t) - 1) - e^{\ln(t)-1+1} = t \ln(t) - 2t.$$

Así que

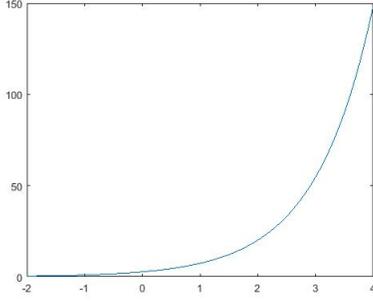
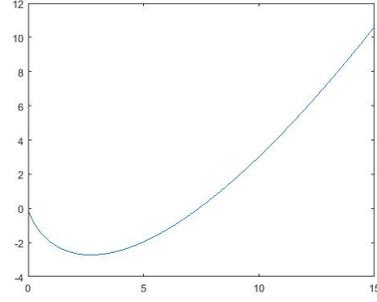
$$\theta^*(t) = t \ln(t) - 2t, \text{ (ver Figura 3.2)}$$

$$\theta^*(e) = e \ln(e) - 2e = e - 2e = -e.$$

Calculando la derivada de  $\theta^*(t)$  tenemos

$$\begin{aligned} (\theta^*)'(t) &= \frac{1}{t}t + \ln(t) - 2, \\ &= \ln(t) + 1, \end{aligned}$$

$$(\theta^*)'(e) = \ln(e) - 1 = 0.$$

Figura 3.1:  $\theta(t) = e^{(t+1)}$ .Figura 3.2:  $\theta^*(t) = t \ln(t) + 2t$ .

### 3.1.2. Shift en función penalidad

Se usará una constante para definir una generalización de la función penalidad la cual sera denominado como shift de función penalidad, esta función será usada en el contexto del método de multiplicación. Se elige  $\tilde{y} \in \mathbb{R}$  así que  $\theta'(\tilde{y}) = c$  con  $c \in \mathbb{R}_{++}$ . por el corolario 1.1 se tiene:

$$(\theta')^{-1} = (\theta^*)'. \quad (3.8)$$

Y así

$$\theta'(\tilde{y}) = c \Leftrightarrow \tilde{y} = (\theta^*)'(c). \quad (3.9)$$

Dado  $p \geq 0, r \in (0, 1]$ , se define la función penalidad generalizada con shift  $P_p$  como:

$$y \in \mathbb{R}^m, \mu \in \mathbb{R}_{++} \mapsto P_p(y, \mu, r, c) = \sum_{i=1}^m P_{p,i}(y_i, \mu_i, r, c). \quad (3.10)$$

Donde  $P_{p,i}(y_i, \mu_i, r, c) = r \frac{\mu_i^p}{c} \left[ \theta \left( \frac{y_i}{y_i^{p-1} r} + \tilde{y} \right) - \theta(\tilde{y}) \right]$  para  $i = 1, \dots, m$  con  $\theta$  satisfaciendo las condiciones 1a)-4a) y  $\theta'(\tilde{y}) = c$ .

Note que:

$$P_{p,i}(0, \mu_i, r, c) = 0 \quad \text{para } i = 1, \dots, m. \quad (3.11)$$

$$(P_{p,i})'_1(y_i, \mu_i, r, c) = \frac{\mu_i}{c} \theta \left( \frac{y_i}{y_i^{p-1} r} + \tilde{y} \right). \quad (3.12)$$

$$(P_{p,i})'(0, \mu_i, r, c) = \mu_i \quad \text{para } i = 1, \dots, m. \quad (3.13)$$

En efecto,

$$\begin{aligned} P_{p,i}(0, \mu_i, r, c) &= r \frac{\mu_i^p}{c} \left[ \theta \left( \frac{0}{y_i^{p-1} r} + \tilde{y} \right) - \theta(\tilde{y}) \right], \\ &= r \frac{\mu_i^p}{c} [\theta(\tilde{y}) - \theta(\tilde{y})], \\ &= r \frac{\mu_i^p}{c} 0, \\ &= 0. \end{aligned}$$

$$\begin{aligned} (P_{p,i})'_1(y_i, \mu_i, r, c) &= r \frac{\mu_i^p}{c} \left[ \theta' \left( \frac{y_i}{y_i^{p-1} r} + \tilde{y} \right) \frac{1}{\mu_i^{p-1} r} \right], \\ &= \frac{\mu_i}{c} \left[ \theta' \left( \frac{y_i}{y_i^{p-1} r} + \tilde{y} \right) \right]. \end{aligned}$$

$$\begin{aligned} (P_{p,i})'_1(0, \mu_i, r, c) &= \frac{\mu_i}{c} \left[ \theta' \left( \frac{0}{y_i^{p-1} r} + \tilde{y} \right) \right], \\ &= \frac{\mu_i}{c} [\theta'(\tilde{y})], \\ &= \frac{\mu_i}{c} c, \quad \text{por (3.9)}, \\ &= \mu_i. \end{aligned}$$

**Proposición 3.1** (ver [5] Proposición 2.2) Sea  $\theta(t)$  una función penalidad satisfaciendo 1a)-4a). Dado  $p \geq 0$ ,  $r \in (0, 1]$ ,  $\tilde{y} \in \mathbb{R}$ ,  $c \in \mathbb{R}_{++}$  considere

$$y \in \mathbb{R}^m, \mu \in \mathbb{R}_{++} \mapsto P_p(y, \mu, r, c) = \sum_{i=1}^m r \frac{\mu_i^p}{c} \left[ \theta \left( \frac{y_i}{y_i^{p-1} r} + \tilde{y} \right) - \theta(\tilde{y}) \right]. \quad (3.14)$$

Cuando  $\theta'(\tilde{y}) = c$  entonces

$$P_{p,\mu,r,c}^*(s) = r \sum_{i=1}^m \left[ \frac{\mu_i^p}{c} \theta \left( \frac{cs_i}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c) (s_i - \mu_i) \right]. \quad (3.15)$$

**Demostración:**

Como

$$P_{p,i}(y_i, \mu_i, r, c) = r \frac{\mu_i^p}{c} \left[ \theta \left( \frac{y_i}{y_i^{p-1} r} + \tilde{y} \right) - \theta(\tilde{y}) \right] \text{ para } i = 1, \dots, m.$$

Donde  $\theta$  es diferenciable en su dominio y  $\theta'(\tilde{y})=c$ .

Sea  $r, c, \mu$  fijas y por la proposición 1.3 se tiene que:

$$\begin{aligned} P_{p,\mu_i,r,c}^*(s_i) &= \left[ r \frac{\mu_i^p}{c} \theta \left( \frac{y_i}{y_i^{p-1} r} + \tilde{y} \right) - \theta(\tilde{y}) \right]^* , \\ &= \left[ r \frac{\mu_i^p}{c} \theta \left( \frac{y_i}{y_i^{p-1} r} \right) + \tilde{y} \right]^* + r \frac{\mu_i^p}{c} \theta(\tilde{y}), \\ &= r \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c r \mu_i^{p-1}}{\mu_i^p r} \right) - \frac{r \mu_i^p}{c} \frac{c s_i}{\mu_i} y_i + r \frac{\mu_i^p}{c} \theta(\tilde{y}), \\ &= r \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - r \mu_i^{p-1} s_i y_i + r \frac{\mu_i^p}{c} \theta(\tilde{y}). \end{aligned}$$

Por el teorema 1.4 si  $\tilde{y} \in \text{dom}(\theta)$  y  $c \in \text{dom}(\theta^*)$  se tiene que  $\theta(\tilde{y}) + \theta^*(c) = c\tilde{y}$  para cada  $i = 1, \dots, m$ .

$$\begin{aligned} P_{p,\mu_i,r,c}^*(s_i) &= r \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - r \mu_i^{p-1} s_i y_i + r \frac{\mu_i^p}{c} [c\tilde{y} - \theta^*(c)], \\ &= r \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - \frac{r}{c} \mu_i^p \theta^*(c) - \mu^{p-1} r \tilde{y} (s_i - \mu_i), \\ &= r \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - \frac{r}{c} \mu_i^p \theta^*(c) - \mu^{p-1} r (\theta^*)'(c) (s_i - \mu_i) , \text{ por (2.9)}. \end{aligned}$$

Así que

$$P_{p,\mu_i,r,c}^*(s_i) = r \sum_{i=1}^m \left[ \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu^{p-1} (\theta^*)'(c) (s_i - \mu_i) \right].$$

### 3.2. La generalización de la casi-distancia

De acuerdo con la definición de 2.1 en [7], dado  $S \subset \mathbb{R}^n$ ;  $d : S \times S \mapsto \mathbb{R}$  es llamada una medida divergente si y solo si:

- i)  $d(x, y) \geq 0$  para todo  $x, y \in S$ .
- ii) si  $\{x^k\} \subset S$  entonces  $\lim_{k \rightarrow \infty} d(x, x^k) = 0$  sii  $\lim_{k \rightarrow \infty} x^k = x$ .
- iii) El conjunto de nivel de  $\Gamma_1(y, \nu) = \{x \in S : d(x, y) \leq \nu\}$  es acotada para toda  $y \in S$  y para toda  $\nu > 0$ .
- iv) El conjunto de nivel de  $\Gamma_2(y, \nu) = \{y \in S : d(x, y) \leq \nu\}$  es acotada para toda  $x \in S$  y para toda  $\nu > 0$ .

Usando (3.15) se puede definir la casi-distancia generalizada como

$$d_{\theta^*}^p(x, y) = r \sum_{i=1}^m \left[ \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu^{p-1} (\theta^*)'(c) (s_i - \mu_i) \right]. \quad (3.16)$$

Note que para  $p = 1$ ,  $c = 1$ ,  $\tilde{y} = 0$  con  $\theta(0) = 0$  se obtiene que

$$p_1^*(s, \mu, r, c) = r d_{\theta^*}(s, \mu) = r \sum_{i=1}^m \mu_i \theta^* \left( \frac{s_i}{\mu_i} \right).$$

En efecto.

Calculando  $\theta^*(1)$

$$\theta^*(1) = \sup_{x \in \mathbb{R}} \{x - \theta(x)\}.$$

Observemos que  $x - \theta(x)$  es continua, cóncava y diferenciable, nos interesa de esta función su máximo a si calculamos su derivada para encontrar su punto critico con lo se obtiene que  $\theta'(x) = 1$ . además por ser  $\theta(x)$  dos veces diferenciables  $\theta^*(x)$  es diferenciable y como  $\theta'(x)$  es distinto de cero podemos aplicar inversa y obtener  $x = (\theta')^{-1}(1) = (\theta^*)'(1) = \tilde{y}$  sustituyendo  $x$  por  $\tilde{y}$  en  $\tilde{y} - \theta(x) = 0 - \theta(\tilde{y})$  así que:

$$\theta^*(1) = -\theta(\tilde{y}) = 0 \text{ por que } \theta(\tilde{y}) = 0. \quad (3.17)$$

$$\begin{aligned}
p_1^*(s, \mu, r, c) &= r \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c) (s_i - \mu_i), \\
&= r \sum_{i=1}^m \frac{\mu_i^1}{1} \theta^* \left( \frac{s_i 1}{\mu_i} \right) - \frac{\mu_i^1}{1} \theta^*(1) - \mu_i^{1-1} (\theta^*)'(c) (s_i - \mu_i), \\
&= r \sum_{i=1}^m \mu_i \theta^* \left( \frac{s_i}{\mu_i} \right) - \mu_i \theta^*(1) - 1 (\theta^*)'(c) (s_i - \mu_i), \\
&= r \sum_{i=1}^m \mu_i \theta^* \left( \frac{s_i}{\mu_i} \right) - \mu_i \theta^*(1) - \tilde{y} (s_i - \mu_i) \text{ por que } (\theta^*)'(c) = \tilde{y}, \\
&= r \sum_{i=1}^m \mu_i \theta^* \left( \frac{s_i}{\mu_i} \right) - \mu_i \theta^*(1), \\
&= r \sum_{i=1}^m \mu_i \theta^* \left( \frac{s_i}{\mu_i} \right) \text{ por (2.17)}.
\end{aligned}$$

$$\begin{aligned}
rd_{\theta^*}(s, \mu) &= r \left( \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c) (s_i - \mu_i) \right), \\
&= p_1^*(s, \mu, r, c).
\end{aligned}$$

Cuando es usado en el método de punto proximal con  $\phi$ -divergencia ver [6], para  $p = 2$ ,  $c = 1$ ,  $\tilde{y} = 0$  con  $\theta(0) = 0$  obtenemos

$$p_1^*(s, \mu, r, c) = rd_{\theta^*}(s, \mu) = r \sum_{i=1}^m \mu_i \theta^* \left( \frac{s_i}{\mu_i} \right)$$

Este resultado se obtiene análogo a el anterior.

**Proposición 3.2** (ver [5] Proposición 5.1)

Dado  $\theta$  satisfaciendo 1b)-4b)  $p \geq 0$ ,  $r \in [0, 1]$  y  $c > 0$ ,

$$d_{\theta^*}^p(x, y) = r \sum_{i=1}^m \left[ \frac{\mu_i^p}{c} \theta^* \left( \frac{s_i c}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c) (s_i - \mu_i) \right],$$

es una medida divergente en  $\mathbb{R}_{++}^n$ .

**Demostración:**

i)

Como  $\theta$  es convexa para toda  $x, y \in \mathbb{R}$  y diferenciable se tiene que

$$\theta^*(z) \geq \theta^*(w) + (\theta^*)'(w)(z - w).$$

Si  $w = c$  y  $z = \frac{cx_i}{y_i}$  para  $i = 1, \dots, m$  con  $x_i \geq 0, y_i \geq 0, \text{dom}\theta = (0, +\infty)$  tenemos

$$\theta^*\left(\frac{cx_i}{y_i}\right) \geq \theta^*(c) + (\theta^*)'(c)\left(\frac{cx_i}{y_i} - c\right),$$

$$\frac{y_i^p}{c} \theta^*\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c} \theta^*(c) - \frac{y_i^p}{c} (\theta^*)'(c) \left(\frac{cx_i}{y_i} - c\right) \geq 0,$$

$$\frac{y_i^p}{c} \theta^*\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c} \theta^*(c) - y_i^{p-1} (\theta^*)'(c) (x_i - y_i) \geq 0.$$

Así

$$\sum_{i=1}^m \left[ \frac{y_i^p}{c} \theta^*\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c} \theta^*(c) - y_i^{p-1} (\theta^*)'(c) (x_i - y_i) \right] \geq 0.$$

Con lo que se tiene que  $d_{\theta^*}^p(x, y) \geq 0$ .

ii)

Sea  $\{x^k\}$  es una sucesión en  $(0, \infty)$  se necesita demostrar que

$$\lim_{k \rightarrow \infty} d(x, x^k) = 0 \Leftrightarrow \lim_{k \rightarrow \infty} x^k = x.$$

$$\begin{aligned} \lim_{k \rightarrow \infty} d(x, x^k) = 0 &\Leftrightarrow \lim_{k \rightarrow \infty} \sum_{i=1}^m \left[ \frac{(x_i^k)^p}{c} \theta^*\left(\frac{cx_i}{x_i^k}\right) - \frac{(x_i^k)^p}{c} \theta^*(c) - (x_i^k)^{p-1} (\theta^*)'(c) (x_i - x_i^k) \right] \\ &= 0, \\ &\Leftrightarrow \lim_{k \rightarrow \infty} \left[ \frac{y_i^p}{c} \theta^*\left(\frac{cx_i}{y_i}\right) - \frac{(x_i^k)^p}{c} \theta^*(c) - (x_i^k)^{p-1} (\theta^*)'(c) (x_i - x_i^k) \right] = 0, \\ &\Leftrightarrow \lim_{k \rightarrow \infty} \theta^*\left(\frac{cx_i}{x_i^k}\right) = \theta^*(c) \text{ para cada } i = 1, \dots, m, \\ &\Leftrightarrow \lim_{k \rightarrow \infty} \left(\frac{cx_i}{x_i^k}\right) = (c) \text{ para cada } i = 1, \dots, m \text{ por ser } \theta \text{ continua,} \\ &\Leftrightarrow \lim_{k \rightarrow \infty} \left(\frac{x_i}{x_i^k}\right) = 1 \text{ para cada } i = 1, \dots, m, \\ &\Leftrightarrow \lim_{k \rightarrow \infty} x_i^k = x_i \text{ para cada } i = 1, \dots, m, \\ &\Leftrightarrow \lim_{k \rightarrow \infty} x^k = x. \end{aligned}$$

iii)

El conjunto de nivel  $\Gamma_1(y, \nu) = \{x \in S : d(x, y) \leq \nu\}$  es acotado para todo  $\nu \in (0, \infty)$ . Suponga que para algún  $y \in (0, \infty)$  y para algún  $\nu \in (0, \infty)$   $\Gamma_1(y, \nu)$  no esta acotado. Entonces existe  $\{x^k\} \in (0, \infty)$  tal que  $x^k \rightarrow \infty$ . Como  $d_{\theta^*}^p(x, y) \leq \nu$  por la definición de  $\Gamma_1$  se tiene que.

$$0 \leq \sum_{i=1}^m \frac{y_i^p}{c} \theta^* \left( \frac{x_i^k c}{y_i} \right) - \frac{y_i^p}{c} \theta^*(c) - y_i^{p-1} (\theta^*)'(c) (x_i^k - y_i) \leq \nu.$$

Para cada  $i = 1, \dots, m$ .

$$\begin{aligned} 0 &\leq \frac{y_i^p}{c} \theta^* \left( \frac{x_i^k c}{y_i} \right) - y_i^{p-1} (\theta^*)'(c) x_i^k \leq \nu + \frac{y_i^p}{c} \theta^*(c) - y_i^{p-1} (\theta^*)'(c) y_i, \\ 0 &\leq \frac{y_i^p}{c} \theta^* \left( \frac{x_i^k c}{y_i} \right) - y_i^{p-1} (\theta^*)'(c) x_i^k \leq \nu + \frac{y_i^p}{c} \theta^*(c) - y_i^p (\theta^*)'(c). \end{aligned}$$

Si  $M_i = \nu + \frac{y_i^p}{c} \theta^*(c) - y_i^p (\theta^*)'(c)$  para  $i = 1, \dots, m$  se tiene:

$$\begin{aligned} 0 &\leq \frac{y_i^p}{c} \theta^* \left( \frac{x_i^k c}{y_i} \right) - y_i^{p-1} (\theta^*)'(c) x_i^k \leq M_i, \\ 0 &\leq \frac{y_i^p}{c} \theta^* \left( \frac{x_i^k c}{y_i} \right) \leq \frac{M_i}{x_i^k} + y_i^{p-1} (\theta^*)'(c). \end{aligned}$$

Por otro lado

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{y_i^p}{c} \left( \frac{\theta^* \left( \frac{x_i^k c}{y_i} \right)}{x_i^k} \right) &= \frac{y_i^p}{c} \lim_{k \rightarrow \infty} \theta^* \left( \frac{x_i^k c}{y_i} \right) x_i^{-k} \text{FI} \left( \frac{\infty}{\infty} \right), \\ &= \frac{y_i^p}{c} \lim_{k \rightarrow \infty} (\theta^*)' \left( \frac{x_i^k c}{y_i} \right) \frac{c}{y_i}, \\ &= y_i^{p-1} \lim_{k \rightarrow \infty} (\theta^*)' \left( \frac{x_i^k c}{y_i} \right), \\ &= y_i^p + \infty. \end{aligned}$$

Si  $c = k (\theta^*)'(c) = 0$  con lo que se tiene una ya que tendríamos que  $+\infty \leq 0$ .

iv)

El conjunto  $\Gamma_2(x, \nu) = \{y \in S : d(x, y) \leq \nu\}$  es acotado. Para todo  $x \in (0, \infty)$  y para todo  $\nu \in (0, \infty)$ .

Suponga que existe una sucesión  $\{y^k\} \subset \Gamma_2(x, \nu)$  tal que para algún  $j$   $\{y_j^k\}$  no es acotada, sin pérdida de generalidad podemos suponer que  $y_j^{k+1} > y_j^k$  para toda  $k$  implicando que  $\frac{x_j^k}{y_j^k} \rightarrow 0$  si  $k \rightarrow \infty$  entonces para  $k$  suficientemente grande  $\frac{x_j}{y_j^k} < 1$  con lo que se tiene que  $\frac{cx_j}{y_j^k} < c$ .

Como  $\{y^k\} \in \Gamma_2(x, \nu)$  se puede calcular la distancia  $d_{\theta^*}^p(x_j, y_j^k)$  con lo que se tiene que:

$$\begin{aligned}
\nu &\geq \frac{(y_j^k)^p}{c} \theta^* \left( \frac{x_j c}{y_j^k} \right) - \frac{(y_j^k)^p}{c} \theta^*(c) - (y_j^k)^{p-1} (\theta^*)'(c) x_j - (y_j^k)^{p-1} (\theta^*)'(c) y_j^k, \\
&\geq \frac{(y_j^k)^p}{c} \theta^* \left( \frac{x_j c}{y_j^k} \right) - \frac{(y_j^k)^p}{c} \theta^*(c) - (y_j^k)^{p-1} (\theta^*)'(c) (x_j - y_j^k), \\
&\geq \frac{(y_j^k)^p}{c} (\theta^*)'(\zeta) \left( \frac{x_j c}{y_j^k} - c \right) - (y_j^k)^{p-1} (\theta^*)'(c) (x_j - y_j^k) \quad T.V.M \quad \frac{x_j c}{y_j^k} < \zeta < c, \\
&\geq \frac{(y_j^k)^p}{c} (\theta^*)'(\zeta) c \left( \frac{x_j}{y_j^k} - 1 \right) - (y_j^k)^{p-1} (\theta^*)'(c) (x_j - y_j^k), \\
&\geq (y_j^k)^p (\theta^*)'(\zeta) \left( \frac{x_j}{y_j^k} - 1 \right) - (y_j^k)^{p-1} (\theta^*)'(c) (x_j - y_j^k), \\
&\geq (y_j^k)^p (\theta^*)'(\zeta) \left( \frac{x_j}{y_j^k} - 1 \right) - (y_j^k)^p (\theta^*)'(c) \left( \frac{x_j}{y_j^k} - 1 \right), \\
&\geq (y_j^k)^p [(\theta^*)'(\zeta) - (\theta^*)'(c)] \left( \frac{x_j}{y_j^k} - 1 \right), \\
&\geq (y_j^k)^p (\theta^*)''(\xi) - \left( \frac{x_j}{y_j^k} - 1 \right) \quad T.V.M \quad \zeta < \xi < c.
\end{aligned}$$

Como  $\{y_j\}$  no es acotado  $\{y_j\} \rightarrow \infty$  esto implica  $(y_j^k)^p (\theta^*)''(\xi) - \left( \frac{x_j}{y_j^k} - 1 \right) \rightarrow \infty$  lo cual es una contradicción ya que  $\nu \in (0, \infty)$ . Así que  $\Gamma_2(x, \nu)$  es acotada.

**Ejemplo 3.2** para  $m = 1$  y  $\theta(t) = e^t$ , (ver Figura 3.3), nosotros tenemos que  $\theta^*(t) = s \ln(s) - s$ ,  $(\theta^*)'(s) = \ln(s)$  y  $d_{\theta^*}^p(s, \mu) = \mu^{p-1} \left( s \ln \left( \frac{s}{\mu} \right) - s + \mu \right)$

En efecto.

$\theta^*(s) \sup_{x \in \mathbb{R}} \{sx - \theta(x)\}$  observe que  $sx - \theta(x)$  es cóncava, continua y diferenciable, como nos interesa es el máximo de esta función calculamos su derivada e igualamos a cero para obtener los puntos críticos así que  $s = e^x$  como  $\theta(t)$  es dos veces diferenciables y distinta de cero podemos aplicar inversa con lo que  $x = \ln(s)$  sustituyendo  $x$  se obtiene que  $\theta^*(s) = s \ln(s) - s$

$$\begin{aligned} (\theta^*)'(s) &= \ln(s) + \frac{s}{s} - 1, \\ &= \ln(s) + 1 - 1, \\ &= \ln(s). \end{aligned}$$

$$\begin{aligned} d_{\theta^*}^p(s, \mu) &= \frac{\mu^p}{c} \left( \frac{sc}{\mu} \ln \left( \frac{sc}{\mu} \right) - \frac{sc}{\mu} \right) - \frac{\mu^p}{c} (c \ln(c) - c) - \mu^{p-1} \ln(c)(s - \mu), \\ &= \mu^{p-1} \left( s \ln \left( \frac{sc}{\mu} \right) - s - \mu(\ln(c) - 1) - \ln(c)(s - \mu) \right), \\ &= \mu^{p-1} \left( s \ln \left( \frac{sc}{\mu} \right) - s - \mu \ln(c) + \mu - \ln(c)s + \ln(c)\mu \right), \\ &= \mu^{p-1} \left( s \ln \left( \frac{sc}{\mu} \right) - s + \mu - \ln(c)s \right), \\ &= \mu^{p-1} \left( s \left( \ln \left( \frac{sc}{\mu} \right) - \ln(c) \right) - s + \mu \right), \\ &= \mu^{p-1} \left( s \ln \left( \frac{s}{\mu} \right) - s + \mu \right). \text{(ver Figura 3.4 para } p=1) \end{aligned}$$

Esta casi-distancia puede ser considerada como la generalización de la casi-distancia Kullback-Leiber la misma puede ser obtenida por  $\theta(t) = e^t - 1$  o de  $\theta(t) = e^{t+k}$  con  $k \in \mathbb{R}$

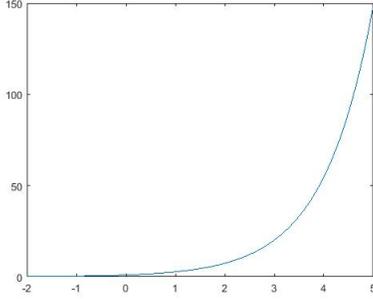


Figura 3.3:  $\theta(t) = e^t$ .

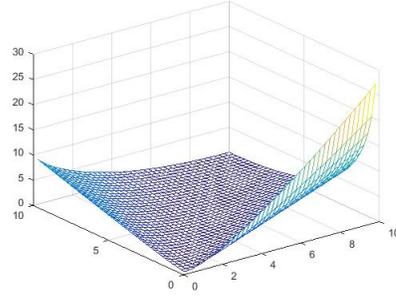


Figura 3.4:  $d_{\theta^*}^1(s, \mu) = s * \ln(s/t) - s + \mu$ .

### 3.2.1. Interpretación geométrica

En esta sección veremos la relación que existe entre la casi-distancia presentada y la distancia de Bregman ya estudiada, además de ver gráficamente como es la función penalidad y su respectiva distancia.

Consideremos la casi-distancia generalizada en (3.16) y dado  $\mu \in \mathbb{R}_{++}^m$  se define la función convexa.

$$h_{\mu} : \mathbb{R}_+^m \rightarrow \mathbb{R} \quad \text{como} \quad h_{\mu} = (s) \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^* \left( \frac{cs_i}{\mu_i} \right).$$

Note que  $h_{\mu}(\mu) = \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^*(c)$  y

$$\nabla h_{\mu}(s) = \left( \mu_1^p (\theta^*)' \left( \frac{cs_1}{\mu_1} \right), \dots, \mu_m^p (\theta^*)' \left( \frac{cs_m}{\mu_m} \right) \right)^t.$$

Así que  $\nabla h_{\mu}(\mu) = (\mu_1^{p-1} \theta^*(c), \dots, \mu_m^{p-1} \theta^*(c))^t$ .

Se tiene que

$$d_{\theta^*}^p(s, \mu) = h_{\mu}(s) - h_{\mu}(\mu) - \nabla h_{\mu}(\mu)^t (s - \mu).$$

Esto es, para cada  $\mu$ ,  $d_{\theta^*}^p(s, \mu)$  es la diferencia de  $h_{\mu}(s)$  y su aproximación lineal en  $s = \mu$ .

Podemos observar que para cada  $\mu > 0$ ,  $h_\mu(s) = \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^* \left( \frac{cs_i}{\mu_i} \right)$  es una función estrictamente convexa.

En efecto.

Tenemos que  $\theta^*$  es convexa por el teorema 1.4 y el producto de una función constante por una función convexa es convexa.

Veamos que para cada  $\mu > 0$ , esto genera una distancia de Bregman  $D_{h_\mu} : \mathbb{R}_+^m \times \mathbb{R}_{++}^m \rightarrow \mathbb{R}$  dada por:

$$D_{h_\mu}(x, y) = h_\mu(x) - h_\mu(y) - \nabla h_\mu(y)^t(x - y).$$

Para probar que esto es una función Bregman se deben verificar las condiciones  $B_1 - B_5$  de la sección 2.4, en efecto.

Para  $B_1$ , como  $\theta^*$  es diferenciable por teorema 1.4.2 y en consecuencia continua  $h_\mu(s)$  es diferenciable y continua ya que esta es producto de  $\theta^*$  y una constante.

Para  $B_2$ , como  $\theta^*$  es estrictamente convexa, el producto de  $\theta^*$  con una constante sigue siendo estrictamente convexa, con lo que se obtiene que  $h_\mu(s)$  es estrictamente convexa.

Para  $B_3$ , como  $h_\mu(s)$  es continua,  $h_\mu(s)$  es semi-continua interiormente como, semi-continuidad inferior implica conjunto de nivel acotados.

Para  $B_4$  se repite las mismas pruebas que en la sección 2.4 para estos respectivos casos. para  $B_5$  podemos decir que esta condicional igual que  $B_4$  no depende de la forma de la función y se satisface.

Con esto se tiene que para cada  $\mu > 0$  la casi distancia generalizada satisface

$$d_{\theta^*}^p(s, \mu) = D_{h_\mu}(s, \mu). \quad (3.18)$$

En efecto.

$$\begin{aligned} D_{h_\mu}(s, \mu) &= h_\mu(s) - h_\mu(\mu) - \nabla h_\mu(\mu)^t (s - \mu), \\ &= \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^* \left( \frac{cs_i}{\mu_i} \right) - \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^*(c) - \sum_{i=1}^m \mu_i^{p-1} (\theta^*)'(c) (s - \mu), \\ &= \sum_{i=1}^m \left( \frac{\mu_i^p}{c} \theta^* \left( \frac{cs_i}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c) (s - \mu) \right), \\ &= d_{\theta^*}^p(s, \mu). \end{aligned}$$

Si se considera una sucesión  $\{\mu^k\}$  en (3.18) se obtiene una sucesión de distancias de Bregman, así que la distancia generalizada para  $\mu = \mu^k$  coincide con  $D_{h_{\mu^k}}(x, y)$  en  $y = \mu^k$ . Esto sugiere la posibilidad del estudio del método con una sucesión de distancias de Bregman dependiendo de un parámetro o variando la función Bregman inducida en cada iteración. Esto no es considerado en esta trabajo.

En la figura 3.5 se muestra la función  $h_\mu(s)$  y su aproximación lineal para los valores  $\mu = 2, \frac{3}{2}, 1, c = \frac{1}{2}, p = \frac{5}{2}$  y  $\theta^*$  como en el ejemplo 3.2 y en la figura 3.6 se muestra la gráfica de  $d_{\theta^*}^p(s, \mu)$  para los mismos valores.

### 3.3. Métodos y teoremas

En esta sección presentaremos el objetivo final de esta trabajo el cual es implementar la casi-distancia propuesta en el método de punto proximal sobre el problema dual lagrangeano y el lagrangeano aumentado del primal y el estudio de convergencia de las sucesiones generadas por el método.

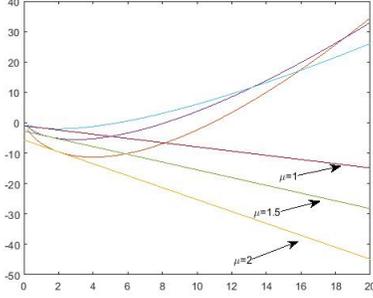


Figura 3.5:

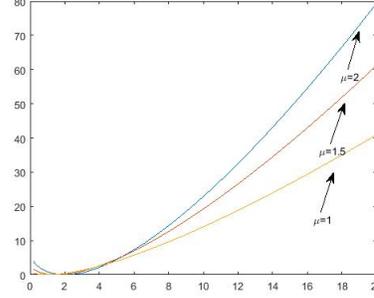


Figura 3.6:

### 3.3.1. Método proximal

En esa sección se aplica el método de punto proximal sobre el dual lagrangeano con la casi-distancia distancia presentada y su convergencia. Consideramos la casi-distancia aplicada directamente sobre el problema dual. El método de punto proximal para resolver (3.2) genera una sucesión  $\{\mu^{k+1}\}$ , tal que  $\{\mu^0\} \in \mathbb{R}_{++}^m$  y

$$\mu^{k+1} = \operatorname{argmin}\{-d(\mu) + r^r d_{\theta^*}^p(\mu^{k+1}, \mu^k)\}. \quad (3.19)$$

Donde  $r^k \in [\tilde{r}, \bar{r}] \subset (0, \infty)$  y  $\theta^*$  definido como en la en la sección(3.1.1) por condiciones de optimalidad se tiene.

$$0 \in \partial[-d(\mu) + r^r d_{\theta^*}^p(\mu^{k+1}, \mu^k)]. \quad (3.20)$$

O equivalentemente

$$-r^r \left( (\mu_1^k)^{p-1} [\theta^{*'} \left( \frac{c\mu_1^{k+1}}{\mu_1^k} \right) - \theta^{*'}(c)], \dots, (\mu_m^k)^{p-1} [\theta^{*'} \left( \frac{c\mu_m^{k+1}}{\mu_m^k} \right) - \theta^{*'}(c)] \right) \in \partial(-d(\mu^{k+1})).$$

**Proposición 3.3** (ver [5] Proposición 3.1)

La sucesión  $\{-d(\mu^k)\}$  es una sucesión convergente no decreciente.

**Demostración:**

Debido a las condiciones de optimalidad en (3.18) se tiene que

$\mu^{k+1} = \operatorname{argmin}\{-d(\mu) + r^r d_{\theta^*}^p(\mu^{k+1}, \mu^k)\}$ , Así que

$$\begin{aligned} -d(\mu^{k+1}) + r^r d_{\theta^*}^p(\mu, \mu^k) &\leq -d(\mu) + r^r d_{\theta^*}^p(\mu^{k+1}, \mu^k), \\ &\leq -d(\mu_k) + r^r d_{\theta^*}^p(\mu^k, \mu^k), \\ &\leq -d(\mu^k). \end{aligned}$$

Por el teorema de dualidad,  $f$  es acotada inferiormente para  $\{-d(\mu^k)\}$ , así que  $\{-d(\mu^k)\}$  es convergente.

**Proposición 3.4** (ver [5] Proposición 3.2)

La sucesión  $\{\mu^k\}$  generada por (3.18) es acotada.

**Demostración:**

Por (H2) ya que  $-d$  es una función propia el conjunto de nivel  $\Lambda = \{\mu \in \mathbb{R}_n^+ / -d(\mu) \leq -d(\mu_0)\}$  es compacto y por al proposición 3.3  $\mu^k \in \Lambda$  para todo  $k$ , así que  $\mu^k$  es acotado.

### 3.3.2. Algoritmo lagrangeano aumentado

En esta sección se aplica el método de punto proximal sobre el lagrangeano aumentado esto debido a que esta función tiene mejores condiciones de convergencia que el lagrangeano estándar .

Considerando el problema (3.1) con hipótesis (H1) y (H2) y con la  $\theta$ -función satisfaciendo las condiciones 1a)-4a). Dado  $p \geq 0$ ,  $r \in (0, 1]$ ,  $\tilde{y} \in \mathbb{R}$ ,  $c \in \mathbb{R}_{++}$  con  $\theta'(\tilde{y}) = c$ , la función lagrangeano aumentado es dada por

$$L_{rc}(x, \mu) = f_0(x) + \sum_{i=1}^m P_{p,i}(f_i(x), \mu_i, r, c).$$

Donde  $P_{p,i}(f_i, \mu_i, r, c) = \frac{\mu_i^p}{c} \left[ \theta \left( \frac{f_i(x)}{\mu_i^{p-1}} + \tilde{y} \right) - \theta(\tilde{y}) \right]$  para  $i = 1, \dots, m$ .

El método de multiplicador a un proximal es dado por:

$$x^{k+1} \in \operatorname{argmin}\{f_0(x) + \sum_{i=1}^m P_{p,i}(f_i(x), \mu_i, r, c)\}. \quad (3.21)$$

$$\mu_i^k = \frac{\mu_i^k}{c} \theta' \left( \frac{f_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + \tilde{y} \right) \text{ para } i = 1, \dots, m \text{ y } r^k \in (0, 1]. \quad (3.22)$$

Observemos que

$$0 \in \partial_x L_{r^k, c}(x^{k+1}, \mu^k) \Leftrightarrow 0 \in \partial_x l(x^{k+1}, \mu^{k+1}). \quad (3.23)$$

Donde  $L$  es la función lagrangeana definida en (3.3) la siguiente proposición muestra que la sucesión definida en (3.22) y en (3.19) son las mismas.

**Proposición 3.5** *Sea  $\{\widehat{\mu}^k\}$  la sucesión generada por (3.19) el cual resuelve el problema dual (3.2) y sea  $\{x_k\}$  y  $\{\mu^k\}$  la sucesión generada por (3.21) y (3.22) el cual resuelve el problema primal (3.1). Si  $\mu^0 = \widehat{\mu}^0$ , entonces  $\forall k > 0 : \mu^k = \widehat{\mu}^k$ .*

La demostración de este teorema se obtiene al aplicar teorema 2.2 directamente sobre este.

### 3.3.3. Resultados de convergencia

Esta sección es inspirada del teorema de convergencia en, [13], y presenta la convergencia del método de punto proximal en el lagrangeano aumentado utilizando la shift en función penalidad.

De (3.16) podemos escribir

$$d(s, \mu) = \sum_{i=1}^m d(s_i, \mu_i). \quad (3.24)$$

Donde  $d(s_i, \mu_i) = \frac{\mu_i^p}{c} \theta^* \left( \frac{cs_i}{\mu_i} \right) - \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c) (s_i - \mu_i)$  para  $i = 1, \dots, m$  y también.

$$d'_1(s_i, \mu_i) = \frac{\partial d(s_i, \mu_i)}{\partial s_i} = \mu_i^{p-1} \left[ (\theta^*)' \left( \frac{cs_i}{\mu_i} \right) - (\theta^*)'(c) \right].$$

**Lema 3.1** Sea  $s, \mu$  números reales positivos con  $s > \mu$  entonces

$$d(s, \mu) \geq \frac{1}{2} \frac{[d'_1(s, \mu)]^2}{cM\mu^{p-2}}$$

donde  $M = \max\{(\theta^*)''(t) : t \geq k = \theta'(0)\}$  y  $c = \theta'(\tilde{y})$ .

**Demostración:**

Considere la función cuadrática

$$q(t) = q(s) + (t - s)d'_1(s, \mu) + \frac{1}{2}(t - s)^2 M\mu^{p-2}.$$

Así

$$\begin{aligned} q'(t) = 0 &\Leftrightarrow d'_1(s, \mu) + (t - s)cM\mu^{p-2} = 0, \\ \Leftrightarrow t &= \frac{-d'_1(s, \mu)}{cM\mu^{p-2}} + \frac{scM\mu^{p-2}}{cM\mu^{p-2}}, \\ \Leftrightarrow t &= \frac{-d'_1(s, \mu)}{cM\mu^{p-2}} + s. \end{aligned}$$

Así

$$t^* = \frac{-d'_1(s, \mu)}{cM\mu^{p-2}} + s. \quad (3.25)$$

Es un minimizador de  $q(\cdot)$ .

Ya que  $s > \mu > 0$ .

$$d'_1(s, \mu) = \mu^{p-1} \left[ (\theta^*)' \left( \frac{cs}{\mu} \right) - (\theta^*)'(c) \right] > 0. \quad (3.26)$$

Entonces de (3.25)  $t^* < s$ , además

$$\begin{aligned} d'_1(s, \mu) &= d'_1(s, \mu) - d'_1(\mu, \mu), \\ &= (s - \mu)d''_1(\tilde{\mu}, \mu) \text{ para algun } \tilde{\mu} \in [\mu, s] \text{ TVM}, \\ &= (s - \mu)\mu^{p-2}(\theta^*)'' \left( \frac{c\tilde{\mu}}{\mu} \right) \text{ derivando}, \\ &\leq (s - \mu)\mu^{p-2}cM. \end{aligned}$$

Así  $\frac{d'_1(s, \mu)}{\mu^{p-2}cM} \leq s - \mu$  y

$$\mu \leq s - \frac{d'_1(s, \mu)}{\mu^{p-2}cM} = t^*. \quad (3.27)$$

Entonces  $\mu \leq t^* < s$ .

Por otro lado para cada  $t \in [\mu, s)$

$$\begin{aligned} d'_1(s, \mu) - d'_1(t, \mu) &= (s - t)d''_1(\bar{t}, \mu) \text{ para algun } t < \bar{t} < s, \\ &= (s - t)\mu^{p-2}(\theta^*)'' \left( \frac{c\bar{t}}{\mu} \right) c, \\ &= (s - t)\mu^{p-2}cM. \end{aligned}$$

Así  $d'_1(s, \mu) \leq d'_1(t, \mu) + (s - t)\mu^{p-2}cM = q'(t) \forall \mu \leq t \leq s$  y

$$d'_1(t, \mu) \geq d'_1(s, \mu) + (t - s)\mu^{p-2}cM = q'(t).$$

Tenemos entonces que

$$d'_1(t, \mu) \geq q'(t) \text{ para todo } \mu \leq t \leq s.$$

Ya que  $\mu \leq t^* < s$  integrando desde  $t^*$  a  $s$  con respecto a  $t$ , se tiene

$$d(s, \mu) - d(t^*, \mu) \geq q(s) - q(t^*).$$

Así  $d(s, \mu) \geq q(s - t^*) + d(t^*, \mu) \geq q(s) - q(t^*)$  y

$$d(s, \mu) \geq q(s) - q(s) - (t^* - s)d'_1(s, \mu) - \frac{1}{2}(t^* - s)^2cM\mu^{p-2}. \quad (3.28)$$

De (3.27) y reemplazando  $t^* - s = \frac{d'_1(s, \mu)}{cM\mu^{p-2}}$  en (3.28) se tiene.

$$d(s, \mu) \geq \frac{[d'_1(s, \mu)]^2}{cM\mu^{p-2}} - \frac{1}{2} \frac{[d'_1(s, \mu)]^2}{cM\mu^{p-2}}, \quad (3.29)$$

$$= \frac{1}{2} \frac{[d'_1(s, \mu)]^2}{cM\mu^{p-2}}. \quad (3.30)$$

Finalmente.

$$d(s, \mu) \geq \frac{1}{2} \frac{[d'_1(s, \mu)]^2}{cM\mu^{p-2}} \quad (3.31)$$

**Proposición 3.6** (ver [5] Proposición 3.5)

(Factibilidad Asintótica) Considere  $p \geq 2$ , la sucesión  $\{[f_i(x^k)]_+\}$  converge a 0 para todo  $i = 1, \dots, m$  donde  $[y] = \max\{0, y\}$ .

**Demostración:**

Suponga se por absurdo que  $\{[f_i(x^k)]_+\} \not\rightarrow 0$  entonces existe  $\{l_k\}$  tal que  $l_k \rightarrow +\infty$  con  $l_k < l_{k+1}$  y  $\exists \varepsilon > 0$  tal que  $f_i(x^{l_k}) > \varepsilon$  para algún  $i_0 = \{1, \dots, m\}$ .

Se tiene que

$$\begin{aligned} d(\mu) &= \inf\{f_0(x) + \sum_{i=1}^n \mu_i f_i(x)\} \leq f_0(x^{k+1}) + \sum_{i=1}^n \mu_i f_i(x^{k+1}), \\ &= f_0(x^{k+1}) + \sum_{i=1}^n \mu_i^{k+1} f_i(x^{k+1}) + \sum_{i=1}^n (\mu_i - \mu^{k+1}) f_i(x^{k+1}), \\ &= d(\mu_i^{k+1}) + \sum_{i=1}^n (\mu_i - \mu^{k+1}) f_i(x^{k+1}) \text{ para todo } i = 1, \dots, m. \end{aligned}$$

De (3.22)

$$\mu_i^{k+1} = \frac{\mu_i^{k+1}}{c} \theta' \left( \frac{f_i(x^{k+1})}{(\mu^k)^{p-1} r^k} + \tilde{y} \right) \text{ para todo } i = 1, \dots, m.$$

Y usando que  $(\theta')^{-1} = (\theta^*)'$  se obtiene

$$r^k (\mu_i^k)^{p-1} \left[ (\theta^*)' \left( \frac{c \mu_i^{k+1}}{\mu_i^k} \right) - (\theta^*)'(c) \right] = f_i(x^{k+1}) \text{ para todo } i = 1, \dots, m. \quad (3.32)$$

Usando la notación en (3.24)

$$d(s, \mu) = \frac{\mu^p}{c} \theta^* \left( \frac{cs_i}{\mu_i} \right) - \frac{\mu^p}{c} \theta^*(c) - \mu^{p-1} (\theta^*)'(c) (s_i - \mu_i) \text{ para } i = 1, \dots, m.$$

Y

$$d'_1(s_i, \mu_i) = \mu_i^{p-1} \left[ (\theta^*)' \left( \frac{cs_i}{\mu_i} \right) - (\theta^*)'(c) \right] \text{ para todo } i = 1, \dots, m. \quad (3.33)$$

De (3.32) (3.33)  $\frac{f_i(x^{k+1})}{r^k} = d_i(\mu_i^{k+1}, \mu_i^k)$  para  $1, \dots, m$  ya que fijo  $(x^{k+1}) > \varepsilon$  entonces  $r^k d'_1(\mu_{i_0}^{k+1}, \mu_{i_0}^k) > \varepsilon > 0$  esto implica  $d'_1(\mu_{i_0}^{k+1}, \mu_{i_0}^k) > 0$  y  $\mu_{i_0}^{k+1} > \mu_{i_0}^k$  y podemos se puede usar la proposición 3.3 tal que  $\mu^{k+1} = \operatorname{argmin}\{-d(\mu) + d_{\theta^*}^p(\mu, \mu^k)\}$  se tiene  $-d(\mu^{k+1}) + d_{\theta^*}^p(\mu, \mu^k) < -d(\mu^k) + d_{\theta^*}^p(\mu^k, \mu^k) = -d(\mu)$  así  $d(\mu_{k+1}) - d(\mu_k) \geq d_{\theta^*}^p(\mu^{k+1}, \mu^k) \geq 0$ .

Entonces

$$d(\mu^{l_{k+1}}) - d(\mu^{l_k}) \geq d_{\theta^*}^p(\mu^{l_{k+1}}, \mu^{l_k}) \quad (3.34)$$

$$= \sum_{i=1}^m d(\mu^{l_{k+1}}, \mu^{l_k}), \quad (3.35)$$

$$\geq d(\mu_{i_0}^{l_{k+1}}, \mu_{i_0}^{l_k}). \quad (3.36)$$

Por proposición 3.4, sea  $\mu$  una cota superior y por el lema 3.1 se tiene

$$d(\mu^{l_{k+1}}) - d(\mu^{l_k}) \geq \frac{1}{2cM} \frac{[d'_1(\mu_{i_0}^{l_{k+1}}, \mu_{i_0}^{l_k})]^2}{(\mu_{i_0}^{p-2})}, \quad (3.37)$$

$$\geq \frac{\varepsilon^2}{2cM \hat{\mu}_{i_0}^{p-2}}, \quad (3.38)$$

$$= \delta, \quad (3.39)$$

$$= 0. \quad (3.40)$$

Para  $p \geq 2$  entonces  $d(\mu^{l_{k+1}}) - d(\mu^{l_k}) \geq \delta > 0$  y tomamos limite cuando  $k \rightarrow \infty$  nosotros tenemos que 0 es mayor que un valor positivo lo cual es una contradicción. Finalmente  $\lim_{k \rightarrow \infty} [f_i(x^k)]_+ = 0$ .

**Afirmación** Para todo  $t$ ,  $t \frac{\theta'(t + \tilde{y})}{c} \geq t$ .

En efecto.

Si.

$$\begin{aligned}
 t > 0 &\Rightarrow t + \tilde{y} > \tilde{y}, \\
 &\Rightarrow \theta'(t + \tilde{y}) > \theta'(\tilde{y}) \text{ convexidad,} \\
 &\Rightarrow \theta'(t + \tilde{y}) > c \theta'(\tilde{y}), \\
 &\Rightarrow \frac{\theta'(t + \tilde{y})}{c} > 1, \\
 &\Rightarrow t \frac{\theta'(t + \tilde{y})}{c} > t.
 \end{aligned}$$

Si.

$$\begin{aligned}
 t > 0 &\Rightarrow t + \tilde{y} < \tilde{y}, \\
 &\Rightarrow \theta'(t + \tilde{y}) < \theta'(\tilde{y}) \text{ convexidad,} \\
 &\Rightarrow \theta'(t + \tilde{y}) < c \theta'(\tilde{y}), \\
 &\Rightarrow \frac{\theta'(t + \tilde{y})}{c} < 1, \\
 &\Rightarrow t \frac{\theta'(t + \tilde{y})}{c} > t \text{ por ser } t < 0.
 \end{aligned}$$

**Proposición 3.7** (ver [5] Proposición 3.6)

Sea  $\{x^k\}$  y  $\{\mu^k\}$  las sucesiones generadas por (3.21) (3.22) entonces  $\lim_{k \rightarrow \infty} \mu_i^k f_i(x^k) = 0$  para cada  $i = 1, \dots, m$ .

**Demostración:**

Por absurdo, supongamos que existe  $i_0 \in \{1, \dots, m\}$ ,  $\varepsilon > 0$  y un conjunto de índices infinito tal que.

$$|\mu_{i_0}^{k_j+1} f_{i_0}(x^{k_j+1})| \geq \varepsilon \text{ para todo } j.$$

Ya que  $\{\mu^k\}$  es acotada, existe  $\hat{\mu} > 0$  tal que:

$$0 < \mu^k \leq \hat{\mu} \text{ para todo } k. \quad (3.41)$$

Por lo tanto.

$$\begin{aligned} |\mu_{i_0}^{k_i+1} f_{i_0}(x^{k_j+1})| &\geq \varepsilon, \\ |f_{i_0}(x^{k_j+1})| &\geq \frac{\varepsilon}{\widehat{\mu}_{i_0}}. \end{aligned}$$

Pero  $\{[f_i(x^{k_j+1})]_+\}$  converge a 0 para toda  $i = 1, \dots, m$  ya que  $f_{i_0}(x^{k_j+1}) \geq \frac{\varepsilon}{\widehat{\mu}_{i_0}}$  es verdad para un conjunto finito de índices  $k_j$  se puede considerar sin pérdida de generalidad.

$$f_i(x^{k_j+1}) \leq -\frac{\varepsilon}{\widehat{\mu}_{i_0}} \text{ para todo } j.$$

Ya que  $(f_1(x^{k+1}), \dots, f_m(x^{k+1}))^t \in \partial d(\mu^{k+1})$  y  $d$  es una función cóncava, entonces

$$\sum_{i=1}^m f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \leq d(\mu_i^{k_j+1}) - d(\mu_i^{k_j}). \quad (3.42)$$

Ya que  $\mu_i^{k+1} = \frac{\mu_i^k}{c} \theta' \left( \frac{f_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + \widetilde{y} \right)$  para  $i = 1, \dots, m$  se tiene

$$f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = \mu_i^{k_j} f_i(x^{k_j+1}) \left( \frac{\mu_i^{k_j+1}}{\mu_i^{k_j}} - 1 \right), \quad (3.43)$$

$$= \mu_i^{k_j} f_i(x^{k_j+1}) \left( \frac{1}{c} \left( \frac{f_i(x^{k_j+1})}{(\mu_i^k)^{p-1} r^k} + \widetilde{y} \right) - 1 \right), \quad (3.44)$$

$$= \frac{\mu_i^{k_j} f_i(x^{k_j+1})}{c} \theta' \left( \frac{f_i(x^{k_j+1})}{(\mu_i^k)^{p-1} r^k} + \widetilde{y} \right) - \mu_i^{k_j} f_i(x^{k_j+1}). \quad (3.45)$$

Ya que  $\theta$  es una función estrictamente convexa,  $\theta''(t) > 0$  para toda  $t \in (-\infty, b)$  entonces la función  $\theta'$  es decreciente y usando (3.45) y la afirmación.

$$f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = \frac{r^k (\mu_i^{k_j})^p}{c} \left[ \frac{f_i(x^{k_j+1})}{(\mu_i^k)^{p-1} r^k} \theta' \left( \frac{f_i(x^{k_j+1})}{(\mu_i^k)^{p-1} r^k} + \widetilde{y} \right) \right] - \mu_i^{k_j} f_i(x^{k_j+1}), \quad (3.46)$$

$$\geq \mu_i^{k_j} f_i(x^{k_j+1}) - \mu_i^{k_j} f_i(x^{k_j+1}), \quad (3.47)$$

$$= 0. \quad (3.48)$$

Ya que  $\{d(\mu^k)\}$  es convergente entonces  $\lim_{j \rightarrow \infty} [d(\mu^{k_j+1}) - d(\mu^{k_j})] = 0$ .

Y así

$$\lim_{j \rightarrow \infty} \sum_{i=1}^m f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \quad (3.49)$$

Ya que  $f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \geq 0$  para cada  $i = 1, \dots, m$ .

Se tiene que  $\lim_{j \rightarrow \infty} f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = 0$  para todo  $i = 1, \dots, m$ .

tambien de (3.45)

$$\lim_{j \rightarrow \infty} \mu_i^{k_j} f_i(x^{k_j+1}) \left[ \frac{1}{c} \theta' \left( \frac{f_i(x^{k_j+1})}{(\mu^{k_j})^{p-1} r^{k_j}} + \tilde{y} \right) - 1 \right] = 0 \text{ para todo } i = 1, \dots, m. \quad (3.50)$$

Por otro lado

$$\frac{f_{i_0}(x^{k_j+1})}{r^{k_j}(\mu_{i_0}^{k_j})^{p-1}} + \tilde{y} \leq \frac{-\epsilon}{r^{k_j}(\mu_{i_0}^{k_j})^{p-1} \hat{\mu}_{i_0}} + \tilde{y} \leq 0 + \tilde{y}.$$

Ya que  $\theta'$  es creciente

$$\theta' \left( \frac{f_{i_0}(x^{k_j+1})}{r^{k_j}(\mu_{i_0}^{k_j})^{p-1}} + \tilde{y} \right) \leq \theta' \left( \frac{-\epsilon}{r^{k_j}(\mu_{i_0}^{k_j})^{p-1} \hat{\mu}_{i_0}} + \tilde{y} \right) < 1. \quad (3.51)$$

De (3.50) y (3.51)

$$\lim_{j \rightarrow +\infty} \mu_{i_0}^{k_j} f_{i_0}(x^{k_j+1}) = 0.$$

Ya que

$$\lim_{j \rightarrow +\infty} f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = 0.$$

Entonces  $\lim_{j \rightarrow +\infty} \mu_{i_0}^{k_j+1} f_{i_0}(x^{k_j+1})$  lo cual es una contradicción.

Consecuentemente  $\mu_i^k f_i(x^k) \rightarrow 0$  para todo  $i = 1, \dots, m$

**Teorema 3.1** (ver [5] Proposición 3.7)

La sucesión  $\{f_0(x^k)\}$  converge a  $\hat{f}$ , cada punto límite de la sucesión  $\{x^k\}$  y  $\{\mu^k\}$  generada por el método (3.21) y (3.22) (o (3.21) y (3.19) son soluciones optimal del problema (3.2) y (3.2) respectivamente).

**Demostración:**

Ya que  $\{x^k\}$  es factible, para todo  $\epsilon > 0$   $f_0(x^k) \geq \hat{f} - \epsilon$  para  $k$  suficientemente grande.

Por la condición (H2) la solución optimal del problema (3.2) es no vacía y compacta se tiene  $\hat{f} = \hat{d}$  donde

$$\hat{f} = \min\{f_0(x) \mid x \in \mathbb{R}^n, f_i(x) \leq 0, i = 1, \dots, m\}$$

Y

$$\hat{d} = \sup\{d(\mu) : \mu \in \mathbb{R}_+^m\}.$$

Además para cada  $\beta < \hat{d}$  el conjunto de nivel  $\{\mu \in \mathbb{R}_+^m : d(\mu) \geq \beta\}$  es compacto.

Pero

$$\hat{f} = \hat{d} \geq d(\mu^k), \quad (3.52)$$

$$= \inf\{l(x, \mu^k)\}, \quad (3.53)$$

$$= l(x, \mu^k), \quad (3.54)$$

$$= f_0(x^k) + \sum_{i=1}^m \mu_i^K f_i(x^k). \quad (3.55)$$

Y acorde con la proposición  $\{\mu_i^K f_i(x^k)\}$  converge a 0 para todo  $i = 1, \dots, m$ , así que para toda  $\epsilon > 0$ .

$$\begin{aligned}
\widehat{f} - \epsilon &\leq f_0(x^k), \\
&\leq \widehat{f} - \sum_{i=1}^m \mu_i^K f_i(x^k), \\
&< \widehat{f} + \epsilon.
\end{aligned}$$

Para  $k$  suficientemente grande así que

$$f_0(x^k) \rightarrow \widehat{f}. \quad (3.56)$$

Por proposición para todo  $\epsilon > 0$  y para  $i = 1, \dots, m$   $f_i(x^k) \leq \epsilon$  para  $k$  suficientemente grande, entonces  $f_0(x^k) \leq \widehat{f} + \epsilon$  y  $f_i(x^k) \leq \epsilon$  para  $i = 1, \dots, m$  y como  $k$  es suficientemente grande, por la condición (H1) y corolario 20 en el conjunto.

$$\{x \in \mathbb{R}^n : f_i(x) \leq \alpha, f_0(x) \leq \beta\}.$$

Es compacto para cada  $\alpha, \beta$  entonces  $\{x^k\}$  es acotada y por la proposición 3.4 la sucesión  $\{\mu^k\}$  es también acotada. Si  $\bar{x}$  es un punto límite de  $\{x^k\}$  entonces por proposición 3.5 y (3.56) es una solución primal optimal y si  $\bar{\mu}$  es un punto límite de  $\{\mu^k\}$  usando (3.55) y proposición 3.7,  $d(\mu) = \widehat{f} = \widehat{d}$ . así  $\bar{\mu}$  es una solución optimal dual.



---

## Conclusión

---

Si bien en este trabajo se realizó un estudio extenso del método de punto proximal, sus ideas fundamentales y la aplicación del método con casi-distancias como distancias de Bregman y  $\varphi$ -divergencia nuestro objetivo fundamental siempre fue la casi-distancia propuesta por **Campos, Castillo y Hernández**, [5].

Dicha casi-distancia generalizada definida sobre el conjugado de una familia de shift constante de funciones penalidad establece una relación entre la familia de penalidades conjugadas y una especie de casi-distancia de Bregman. Esta distancia generalizada tiene como casos particulares la  $\varphi$ -divergencia y las funciones kernel homogéneos de segundo orden. Antes de esta generalización la única relación que se tenía entre la distancia de Bregman y la  $\varphi$ -divergencia era la casi-distancia de Kullback-Leibler, con lo que esta generalización puede ser considerada como la casi-distancia generalizada de Kullback-Leibler.

Debemos mencionar que en este trabajo se obtienen resultados de convergencia para el método de punto proximal sobre el problema dual con la casi-distancia generalizada y

el lagrangeano aumentado para el caso en que  $p \geq 2$  que dando abierto el caso  $p < 0$  y  $1 < p < 2$ .

---

## Bibliografía

---

- [1] APOSTOL, T. Análisis matemático. *Editorial Reverté* (2004).
- [2] AUSLENDER, A., TEBoulLE, M., AND BEN-TIBA, S. Interior proximal and multiplier methods based on second order homogeneous kernels. *Mathematics of Operations Research* 3 (1999), 645–668.
- [3] BERTSEKAS, D. Constrained optimization and lagrange multiplier methods. *Academic Press, New York* (1982).
- [4] ECKSTEIN, J. Nonlinear proximal point algorithms using bregman functions with applications to convex programming. *Mathematics of Operations Research* 18 (1993), 202–226.
- [5] E.HERNANDEZ, J.CAMPOS, AND CASTILLO, R. A. A generalized like-distance in convex programming. *International Mathematical Forum* 2 (2007), 1811–1830.
- [6] IUSEM, A. Métodos de ponto proximal em otimização. *20 Coloquio Brasileiro de Matemática* (1995).

- [7] IUSEM, A. Augmented lagrangian methods and proximal point methods for convex optimization. *Investigacion Operativa* 8 (1997), 11–50.
- [8] IUSEM, A. Steepest descent methods with generalized distances for constrained optimization. *Acta Applicande Mathematica* (1997).
- [9] IUSEM, A., TEBoulLE, M., AND SVAITER, B. Entropy-like proximal methods in convex programming. *Mathematics of Operations Research* 4 (1994), 790–814.
- [10] LEMARECHAL, C., AND HIRIART-URRUTY. Fundamentals of convex analysis. *Springer Verlag* (2004).
- [11] PATRIKSOSON, N. A. A. E. M. An introction to continuous optimization. *Professional Plublishing* (2005).
- [12] ROCKAFELLAR, R. Convex analysis. *Princeton University Press, Princeton* (1970).
- [13] TAL, A. M. B., AND ZIBULEVSKY. Penalty/barrier multiplier methods for convex programming problems. *SIAM Journal on Optimization* 7 (1997), 447–336.