

UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”
DECANATO DE CIENCIAS
DEPARTAMENTO DE INV. DE OPERACIONES Y ESTADISTICA

Una Casi-Distancia Generalizada en Programación Convexa.

Caso $0 < p \leq 1$

Trabajo de Ascenso presentado para optar
a la Categoría de Asistente.

Autor: Lcda. Raquel S. Quintana C.

Tutor: Dr. Eibar R. Hernández R.

Barquisimeto, octubre de 2011

Dedicado al Rey de Reyes nuestro Señor Jesucristo.

Porque el Señor da la sabiduría,
Y de su boca viene el conocimiento
y la inteligencia.

Prov. 2:6

Agradecimientos

A Jesucristo por brindarme la oportunidad de ser llamada hija de Dios; A mi padre Germán Saúl por enseñarme la importancia de la perseverancia, a mi madre Loida Esmirna por demostrarme su gran amor cada día.

A mí Pastor Roberto Bacino y todos mis hermanos en Cristo por enseñarme la sana doctrina.

A mi esposo Humberto Bacino por enseñarme que nada es imposible cuando se tiene a Cristo en el corazón.

A mis hijos Anyely y Abraham por mostrarme la importancia de una buena educación.

A mi hermana Clavel, por su incasable apoyo, sus constantes correcciones a lo largo de toda mi vida.

A mis hermanas Claravel y Rosana, por sus incansables consejos en los momentos de debilidades.

A Jorge Moreno por su amistad y su gran apoyo durante este estudio.

A mi tutor el Dr. Eibar Hernandez que me ayudo con sus conocimientos y su tiempo para guiarme en la elaboración de este trabajo.

A todos y cada uno de mis Profesores que me han impartido una buena enseñanza a lo largo de mi vida .

Resumen

Se considera una casi-distancia generalizada, que contiene como casos degenerados, φ -divergencias y casi-distancias con núcleos homogéneos de segundo orden. La motivación principal es estudiar caminos para las casi-distancias generalizadas y obtener la convergencia para valores de p que permanecían abiertos como el caso $0 < p \leq 1$, usando las técnicas estudiadas por Iusem, Svatir y Teboulle observando que dichas técnicas no resultan conveniente para obtener resultados de convergencia para todo $c > 0$, por problemas de no acotabilidad, pero para el caso particular de $c = \bar{k}$ se logra obtener dichos resultados, lo que sugiere que el shift constante a considerar debe ser igual al mínimo de la función de penalidad trasladada.

Índice general

Agradecimientos	4
Resumen	5
Introducción	9
1. Preliminares.	11
1.1. Conceptos y Teorema Básicos	11
1.2. Optimalidad y dualidad en programación no lineal	23
2. Métodos de Punto Proximal y de Lagrangeano Aumentado.	28
2.1. El concepto de regularización	28
2.2. Método de Punto Proximal	30
2.3. Operadores Monótonos Maximales	36
2.4. Lagrangeano Aumentado.	41
2.5. Penalización	43
2.6. Distancia de Bregman y Método de Punto Proximal	44

2.7.	φ -Divergencia y Método de Punto Proximal	54
3.	Una Casi-distancia Generalizada en los casos $0 < p < 2$	63
3.1.	Introducción a la casi-distancia generalizada.	63
3.2.	Familia de Penalidad propuesta.	65
3.3.	Algoritmo proximal con una Casi-distancia Generalizada.	
	Caso $0 < p < 2$	69
3.4.	Análisis de convergencia.	72
	3.4.1. Resultados Preliminares.	72
	3.4.2. Casí-Fejer convergencia.	79
	3.4.3. Resultados de Convergencia.	81
3.5.	Convergencia desde el punto de vista Primal	93
4.	Conclusiones.	98

Índice de figuras

3.1. Función θ	66
3.2. Función θ^*	66
3.3. Función $\varphi(t) \leq \widehat{\varphi}(t)$	69
3.4. $c = \bar{k} = 20, p = 1,6, d_{\widehat{\varphi}} \geq d_{\varphi}$	74
3.5. Contraejemplo de Fejér convergencia	80

Introducción

En el marco de la programación convexa, dado el problema

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

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:

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

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, via dualidad de Fenchel, han sido estudiados ampliamente en los últimos años. Estos métodos han sido estudiados básicamente desde dos puntos de vista, el punto de

vista primal-dual [2],[3] y [15] y el punto de vista dual-primal, [1],[6] y [11].

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 [19] se expresa que las casi-distancias asociadas a φ – *divergencias*, núcleos cuadráticos y distancias de Bregman todas son casos particulares de una casi-distancia generalizada. En el siguiente trabajo se encuentra teoría para desarrollar una casi-distancia generalizada para valores de $0 < p \leq 1$ que contiene como caso particular φ – *divergencias*, la cual permite generalizar la casi-distancia entrópica Kullback-Leibler, estudiando una familia de funciones, que necesariamente no 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 originales de este trabajo, ofreciendo una nueva familia de funciones de penalidad y se construye la casi-distancia generalizada para valores de $0 < p \leq 1$.

Además se incluye un método y teoremas de convergencia en los espacios dual y primal, siguiendo los pasos de [11].

Capítulo 1

Preliminares.

1.1. Conceptos y Teorema Básicos

Dado que el problema que se estudiará más adelante corresponde a un problema de programación convexa, se enunciará algunas definiciones y teoremas fundamentales sobre el análisis convexo, con el fin de facilitar la comprensión y el análisis de los temas que serán desarrollados en los próximos capítulos. El propósito es establecer una base teórica suficiente para describir los métodos de punto proximal y los de multiplicadores. En particular se estudiará el tema de funciones convexas, diferenciabilidad y subdiferenciabilidad, así como las propiedades de su conjugada. Cada uno pueden encontrarse en [2],[13] y [16]. A lo largo de éste, \mathbb{R} denotará el sistema de los números reales y \mathbb{R}^n el espacio vectorial usual de n-uplas de la forma $x = (x_1, \dots, x_n)^t$, donde cada $x_i \in \mathbb{R}$, para $i = 1, \dots, n$.

El producto interno de dos vectores $x = (x_1, \dots, x_n)$ y $y = (y_1, \dots, y_n)$ en \mathbb{R}^n será representado por:

$$\langle x, y \rangle = x_1 y_1 + \dots + x_n y_n = x^t y.$$

Si $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ es una función diferenciable, se denotará por $\nabla_1 f(x, y)$ y $\nabla_2 f(x, y)$ a los gradientes de f respecto a la primera y segunda variable respectivamente.

Los ortantes positivo y no negativo serán representados por los conjuntos

$$\mathbb{R}_{++}^n = \{x \in \mathbb{R}^n : x > 0\} \quad \text{y} \quad \mathbb{R}_+^n = \{x \in \mathbb{R}^n : x \geq 0\},$$

donde $x > 0$ denotará que $x_i > 0$, para todo $i \in \{1, \dots, n\}$ y $x \geq 0$ denotará que $x_i \geq 0$, para todo $i \in \{1, \dots, n\}$.

Definición 1.1.1

Sea C un conjunto convexo y no-vacío en \mathbb{R}^n . Una función $f : C \rightarrow \mathbb{R}$ se dice convexa sobre C si y sólo si para cada par $(x, y) \in C \times C$ y para cada $\lambda \in (0, 1)$ se tiene que:

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

Se dice que f es estrictamente convexa sobre C , cuando la desigualdad anterior es estricta para $x \neq y$.

Definición 1.1.2

Una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ no idénticamente igual a $+\infty$ se dice que es convexa, cuando para todo par $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ y para todo $\lambda \in (0, 1)$ se cumple que:

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

Obsérvese que la desigualdad anterior es en $\mathbb{R} \cup \{+\infty\}$ y se denotará a la clase de tales funciones por $Conv\mathbb{R}^n$.

Definición 1.1.3

El dominio efectivo de $f \in Conv\mathbb{R}^n$ es el conjunto no-vacío

$$dom f = \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

Recuérdese que el gráfico de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es el conjunto:

$$\{(x, f(x)) : x \in \mathbb{R}^n\},$$

lo cual nos permite visualizar la siguiente definición.

Definición 1.1.4

Dada $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función no idénticamente igual a $+\infty$, el epígrado de f es el conjunto no vacío dado por:

$$\text{epi}(f) = \left\{ (x, y) \in \mathbb{R}^n \times \mathbb{R} : y \geq f(x) \right\}$$

El epígrafo estricto es definido de manera similar, reemplazando " \geq " por " $>$ ".

Proposición 1.1.1 (Ver Cap. IV, Prop. 1.1.6, [13])

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función no idénticamente igual a $+\infty$.

Las siguientes propiedades son equivalentes:

- (i) $f \in \text{Conv}\mathbb{R}^n$.
- (ii) $\text{epi}(f)$ es un conjunto convexo en $\mathbb{R}^n \times \mathbb{R}$.
- (iii) " $\text{epi}(f)$ estricto" es un conjunto convexo en $\mathbb{R}^n \times \mathbb{R}$.

Definición 1.1.5

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función. Se llaman conjuntos de sub-nivel de f a los siguientes, para cada $r \in \mathbb{R}$;

$$S_r(f) := \{x \in \mathbb{R}^n : f(x) \leq r\}.$$

Teorema 1.1.1 (ver Cap. IV, Teo. 1.1.8, [13])

Sea $f \in \text{Conv}\mathbb{R}^n$; entonces para cualquier colección $\{x_1, \dots, x_k\}$ de puntos en $\text{dom}f$ y para cualquier $\lambda = (\lambda_1, \dots, \lambda_k)$ en el simplex unitario de \mathbb{R}^k , dado por: $\{\lambda = (\lambda_1, \dots, \lambda_k) : 0 \leq \sum_{i=1}^k \lambda_i \leq 1\}$ se cumple:

$$f\left(\sum_{i=1}^k \lambda_i x_i\right) \leq \sum_{i=1}^k \lambda_i f(x_i).$$

Definición 1.1.6

Una función convexa $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ se dice propia si y sólo si su epígrafo es no vacío y no contiene líneas verticales, es decir, si $f(x) < +\infty$ para al menos un x y $f(x) > -\infty$ para cada x .

Por tanto, f es una función convexa propia si el conjunto convexo $C = \text{dom}f$ es no vacío y la restricción de f a C es finita.

Observe que si $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ es convexa, no idénticamente igual a $+\infty$, entonces f es propia.

Teorema 1.1.2 (Ver Teorema 1.1, [16])

Los subespacios de \mathbb{R}^n son conjuntos afines que contienen el origen.

Teorema 1.1.3 (Ver Teo. 1.5, [16])

Las transformaciones afines de \mathbb{R}^n en \mathbb{R}^m son funciones T de la forma $Tx = Ax + a$, donde Ax es una transformación lineal y $a \in \mathbb{R}^m$.

Ahora bien, el epígrafo de una función lineal es caracterizado por algún $s \in \mathbb{R}^n$ y está formado por aquellos puntos $(x, r) \in \mathbb{R}^n \times \mathbb{R}$ tales que $r \geq \langle s, x \rangle$.

Algunas notaciones que se utilizarán son:

- La cápsula afín de $dom f$ se denotará por $aff(dom f)$ y se define como la intersección de todos los conjuntos afines que contienen a $dom f$ [16].
- El interior relativo del epígrafo de f se denotará por

$$ri(epi(f)) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : x \in ri(dom f), r > f(x)\}.$$

- El interior relativo del dominio efectivo de f se denotará por $ri(dom f)$ y es la proyección sobre \mathbb{R}^n de $ri(epi(f))$.

Definición 1.1.7

Una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ es *semicontinua inferior* si para cada $x \in \mathbb{R}^n$ se cumple que:

$$\liminf_{y \rightarrow x} f(y) \geq f(x).$$

Observe que esta desigualdad se debe cumplir en $\mathbb{R} \cup \{+\infty\}$. Mediante la siguiente proposición se puede apreciar su aspecto geométrico.

Proposición 1.1.2 (Ver Cap. IV, Prop. 1.2.2, [13])

Para $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ las siguientes propiedades son equivalentes:

- (i) f es *semicontinua inferior* en \mathbb{R}^n .
- (ii) $epi(f)$ es un conjunto cerrado en $\mathbb{R}^n \times \mathbb{R}$.
- (iii) Para todo $r \in \mathbb{R} : S_r(f)$ es cerrado (posiblemente vacío).

Definición 1.1.8

La función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ se dice *cerrada* si y sólo si es *semi-continua inferior*, o su epígrafo es cerrado, o sus conjuntos de sub-nivel son cerrados.

Definición 1.1.9

La clausura de una función $f : \mathbb{R}^n \rightarrow R \cup \{+\infty\}$ es la función $clf : \mathbb{R}^n \rightarrow R \cup \{+\infty\}$ definida por:

$$clf(x) := \liminf_{y \rightarrow x} f(y)$$

para todo $x \in \mathbb{R}^n$ o equivalentemente, se define como aquella función que cumple lo siguiente:

$$epi(clf) = cl(epi(f)).$$

Proposición 1.1.3 (Ver Cap. IV, Prop. 1.2.6, [13])

Para $f \in Conv\mathbb{R}^n$ se tiene que:

$$clf \in Conv\mathbb{R}^n;$$

y clf y f coinciden en $ri(dom.f)$.

El conjunto de las funciones convexas cerradas sobre \mathbb{R}^n será denotado por $\overline{Conv}\mathbb{R}^n$.

Veamos algunas proposiciones sobre operaciones funcionales que preservan la convexidad.

Proposición 1.1.4 (Ver Cap. IV, Prop. 2.1.1, [13])

Sean $f_1, \dots, f_n \in Conv\mathbb{R}^n$ [respectivamente en $\overline{Conv}\mathbb{R}^n$], sean t_1, \dots, t_n reales positivos y supóngase que todas las funciones son propias, entonces la función:

$$f := \sum_{j=1}^m t_j f_j$$

está en $Conv\mathbb{R}^n$ [Resp. en $\overline{Conv}\mathbb{R}^n$].

Definición 1.1.10

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función propia no necesariamente convexa, se define la función conjugada f^* mediante:

$$f^*(s) = \sup\{\langle s, x \rangle - f(x) : x \in \text{dom}f\}$$

para todo $s \in \mathbb{R}^n$.

Proposición 1.1.5 (Ver Cap. IV, Prop. 2.1.5, [13])

Sea $f \in \text{Conv}\mathbb{R}^n$ [resp. en $\overline{\text{Conv}}\mathbb{R}^n$] y sea $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$ una transformación afín tal que $\text{Im}A \cap \text{dom}f \neq \emptyset$, donde $\text{Im}A$ es la imagen de A . Entonces la función:

$$f \circ A : \mathbb{R}^m \ni x \mapsto (f \circ A)(x) = f(A(x))$$

está en $\text{Conv}\mathbb{R}^m$ [resp. en $\overline{\text{Conv}}\mathbb{R}^m$].

Proposición 1.1.6 (Ver Cap. IV, Prop. 2.1.8, [13])

Sean $f, g \in \text{Conv}\mathbb{R}^n$ con g creciente. Supóngase que $\exists x_0 \in \mathbb{R}^n / f(x_0) \in \text{dom}g$ y hágase $g(+\infty) := +\infty$. Entonces la función compuesta $g \circ f : x \mapsto g(f(x))$ está en $\text{Conv}\mathbb{R}^n$.

Definición 1.1.11

Sea C un conjunto convexo, cerrado y no-vacío. Para cada $x \in C$, se define el cono asintótico en x como sigue:

$$C_\infty(x) := \{d \in \mathbb{R}^n : x + td \in C \text{ para todo } t > 0\}.$$

$C_\infty(x)$ puede ser visto como el conjunto de todas las direcciones desde las cuales se puede ir en línea recta desde x hacia el infinito sin salir de C .

Proposición 1.1.7 (Ver Cap. IV, Prop. 2.2.1, [13])

El cono convexo cerrado $C_\infty(x)$ no depende de $x \in C$. Esto permite escribir C_∞ en lugar de $C_\infty(x)$.

Proposición 1.1.8 (Ver Cap. IV, Prop. 2.2.3, [13])

Un conjunto C convexo y cerrado es compacto si y sólo si $C_\infty = \{0\}$.

Proposición 1.1.9 (Ver Cap. IV, Prop. 3.2.1, [13])

Sea $f \in \overline{\text{Conv}}\mathbb{R}^n$, entonces el cono asintótico del epígrafo de f , $(\text{epi}(f))_\infty$, es el epígrafo de la función $f'_\infty \in \text{Conv}\mathbb{R}^n$ definida por:

$$d \in \mathbb{R}^n \mapsto f'_\infty(d) := \sup_{t>0} \frac{f(x_0 + td) - f(x_0)}{t} = \lim_{t \rightarrow +\infty} \frac{f(x_0 + td) - f(x_0)}{t},$$

donde x_0 es arbitrario en el dominio de f .

Definición 1.1.12

A esta función f'_∞ se le llama función de recesión.

Proposición 1.1.10 (Ver Cap. IV, Prop. 3.2.5, [13])

Sea $f \in \overline{\text{Conv}}\mathbb{R}^n$. Todos los conjuntos de sub-nivel de f , no-vacío, tiene el mismo cono asintótico, el cual es el conjunto de sub-nivel de f'_∞ en el nivel 0.

Es decir, para todo $r \in \mathbb{R}$ con $S_r(f) \neq \phi$, se tiene que:

$$[S_r(f)]_\infty = \{d \in \mathbb{R}^n : f'_\infty(d) \leq 0\}$$

En particular, las siguientes afirmaciones son equivalentes:

- (i) Existe $r \in \mathbb{R}$ para el cual $S_r(f)$ es no-vacío y compacto.

(ii) Todos los subconjuntos de sub-nivel de f son compactos.

(iii) $f'_\infty(d) > 0$ para todo d no nulo en \mathbb{R}^n

Definición 1.1.13

Las funciones $f \in \overline{Conv}\mathbb{R}^n$ que satisfacen (i), (ii), (iii) en la proposición anterior se llama 0-coercivas. Equivalentemente las funciones 0-coercivas son aquellas que crecen al infinito, es decir:

$$f(x) \rightarrow +\infty \quad \text{cuando} \quad \|x\| \rightarrow +\infty,$$

además, las funciones 0-coercivas, convexas y cerradas alcanzan su mínimo sobre \mathbb{R}^n .

Un importante caso particular es cuando se tiene que:

$$\frac{f(x)}{\|x\|} \rightarrow +\infty \quad \text{cuando} \quad \|x\| \rightarrow +\infty.$$

En este caso f crece al infinito más rápido que cualquier función afín y tales funciones son llamadas 1-coercivas. En algunos textos no se hace distinción entre 0-coercividad y 1-coercividad y a ambas las denominan simplemente coercividad. En forma análoga se puede definir coercividad lateral para el caso en que haya coercividad por un solo lado.

Definición 1.1.14

Una dirección $d \neq 0$ es una dirección de recesión de f si y solo si $f'_\infty(d) \leq 0$.

Teorema 1.1.4 (Ver Cap. IV, Teo. 4.1.4, [13])

Sea f una función diferenciable sobre un conjunto abierto $\Omega \subset \mathbb{R}^n$ y sea C un subconjunto convexo de Ω . Entonces:

(i) f es convexa en C si y sólo si:

$$f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle \quad \text{para todo } (x_0, x) \in C \times C.$$

(ii) f es estrictamente convexa en C si y sólo si la desigualdad anterior es estricta cuando $x \neq x_0$.

Teorema 1.1.5 (Ver Cap. IV, Teo. 4.3.1, [13])

Sea f dos veces diferenciable sobre un conjunto convexo abierto $\Omega \subset \mathbb{R}^n$. Entonces:

(i) f es convexa sobre Ω si y sólo si el Hessiano $\nabla^2 f(x_0)$ es semidefinido positivo para todo $x_0 \in \Omega$.

(ii) Si el Hessiano $\nabla^2 f(x_0)$ es definido positivo para todo $x_0 \in \Omega$, entonces f es estrictamente convexa en Ω .

Definición 1.1.15

Una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ es suave si y sólo si es finita y diferenciable en todo \mathbb{R}^n .

Definición 1.1.16

Una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ convexa y propia es esencialmente suave, si satisface las siguientes condiciones:

(i) $\text{int}(\text{dom} f) \neq \emptyset$.

(ii) f es diferenciable en $\text{int}(\text{dom} f)$.

(iii) $\lim_{k \rightarrow +\infty} \|\nabla f(x_k)\| = +\infty$ donde $\{x_k\}$ es una sucesión en $\text{int}(\text{dom} f)$, que converge a un punto x de la frontera de $\text{int}(\text{dom} f)$.

Obsérvese que si f es diferenciable en \mathbb{R}^n , entonces f es esencialmente suave, pues $\text{int}(\text{dom}f) = \mathbb{R}^n$ y (iii) se cumple por descarte, pues la frontera de \mathbb{R}^n es vacía.

Definición 1.1.17

Sea $f \in \text{Conv}\mathbb{R}$, sea x_0 un punto del interior de su dominio, llamamos derivada por la izquierda y derivada por la derecha, respectivamente a las siguientes:

$$D_-f(x_0) := \lim_{x \uparrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \sup_{x < x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

$$D_+f(x_0) := \lim_{x \downarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \inf_{x > x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Teorema 1.1.6 (Ver Cap. IV, Teo. 4.1.1, [13])

Sea $f \in \text{Conv}\mathbb{R}$, sea x_0 un punto del interior de su dominio, entonces f admite derivadas por la izquierda y por la derecha finitas y satisfacen:

$$D_-f(x_0) \leq D_+f(x_0).$$

Definición 1.1.18

Sean $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ una función convexa. Sean x y d fijos en \mathbb{R}^n y consideremos el cociente incremental de f en x en la dirección d .

$$q(t) := \frac{f(x + td) - f(x)}{t} \quad \text{para } t > 0,$$

se define la función uno-dimensional $\mapsto \varphi(t) := f(x + td)$ entonces $f'(x, d) = D_+\varphi(0)$.

por otro lado:

$$f'(x, -d) = \lim_{x \downarrow x_0} \frac{f(x - td) - f(x)}{t} = \lim_{x \downarrow x_0} \frac{f(x + td) - f(x)}{-t}$$

luego $f'(x, -d) = -D_-\varphi(0)$.

Definición 1.1.19

El subdiferencial $\partial f(x)$ de f en x es el conjunto convexo, compacto y no vacío de \mathbb{R}^n definido como:

$$\partial f(x) := \{s \in \mathbb{R}^n : \langle s, d \rangle \leq f'(x, d) \text{ para todo } d \in \mathbb{R}^n\}$$

Un vector $s \in \partial f(x)$ es llamado subgradiente de f en x .

Definición 1.1.20

El conjunto subdiferencial de f en x_0 , denotado por $\partial f(x_0)$, es el siguiente:

$$\partial f(x_0) := \{s \in \mathbb{R}^n : f(x) \geq f(x_0) + \langle s, x - x_0 \rangle \text{ para todo } x \in \mathbb{R}^n\}.$$

Teorema 1.1.7 (Ver Cap. IV, Teo. 4.3.1, [13])

Las definiciones 1.1.21 y 1.1.22 son equivalentes: La definición 1.1.22 significa que los elementos de $\partial f(x_0)$ son los vectores directores de los hiperplanos que pasan por $(x_0, f(x_0)) \in \mathbb{R}^n \times \mathbb{R}$.

Teorema 1.1.8 (Ver Cap. IV, Teo. 2.2.1, [13])

Para $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexa, las siguientes tres propiedades son equivalentes:

- (i) f es minimizada en x sobre \mathbb{R}^n .
- (ii) $0 \in \partial f(x)$.
- (iii) $f'(x, d) \geq 0$ para todo $d \in \mathbb{R}^n$.

Una regla de cálculo con subdiferenciales es dada por el siguiente teorema:

Teorema 1.1.9 (Ver Cap. IV, Teo. 4.4.1, [13])

Sean f_1, f_2 dos funciones convexas de $\mathbb{R}^n \rightarrow \mathbb{R}$ y t_1, t_2 positivos, entonces:

$$\partial(t_1 f_1 + t_2 f_2)(x) = t_1 \partial f_1(x) + t_2 \partial f_2(x), \quad \text{para todo } x \in \mathbb{R}^n.$$

1.2. Optimalidad y dualidad en programación no lineal

En esta sección se estudiarán algunos conceptos básicos sobre las condiciones necesarias de optimalidad en programación no lineal, describiendo las condiciones de Karush-Kuhn-Tucker y aquellos resultados que muestran que estas condiciones son también suficientes para problemas convexos. Además se ofrecerá una breve descripción de la teoría de dualidad.

El problema de programación no lineal se puede formular como:

$$\min\{f(x) : x \in R^n\}, \quad (1.1)$$

sujeto a,

$$h(x) = 0, \quad (1.2)$$

$$g(x) \leq 0. \quad (1.3)$$

donde $x = (x_1, \dots, x_n)^t$ es el vector de las variables de decisión, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es la función objetivo, 1.2 y 1.3 son respectivamente las restricciones de igualdad y de desigualdad, donde al menos una de las restricciones es no lineal. A cualquier vector que satisfaga las restricciones se la llamará solución factible y el conjunto de todas las soluciones factibles se la llamará región factible.

El problema anterior en el caso en que todas las funciones son convexas se la llama programa convexo.

Es bien conocido que si S es un conjunto cerrado, acotado y no vacío de \mathbb{R}^n , es decir, compacto y $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continua, entonces el problema de hallar

$\min\{f(x) : x \in S\}$, posee al menos una solución óptima. Por otra parte, el conjunto definido por las restricciones bien pudiera no ser acotado, en cuyo caso se puede garantizar la existencia de soluciones óptimas utilizando hipótesis adicionales convenientes.

Uno de los resultados teóricos más importantes en la programación no lineal es el que lleva a las llamadas condiciones de Karush-Kuhn-Tucker. Estas condiciones deben ser satisfechas por la solución óptima de cualquier problema lineal o no lineal.

Definición 1.2.1 *Condiciones de Karush-Kuhn-Tucker (KKT).*

El vector $\bar{x} \in \mathbb{R}^n$ satisface las condiciones KKT para el problema no lineal (en el caso diferenciable) (1.1), (1.2), (1.3), si existe un par de vectores $\lambda \in \mathbb{R}^\ell$ y $\mu \in \mathbb{R}^m$ tales que:

$$\nabla f(\bar{x}) + \sum_{k=1}^{\ell} \lambda_k \nabla h_k(\bar{x}) + \sum_{j=1}^m \mu_j \nabla g_j(\bar{x}) = 0, \quad (1.4)$$

$$h_k(\bar{x}) = 0, \quad (1.5)$$

$$g_j(\bar{x}) \leq 0, \quad (1.6)$$

$$\mu_j g_j(\bar{x}) = 0, \quad (1.7)$$

$$\mu_j \geq 0. \quad (1.8)$$

A los vectores λ y μ se les llama multiplicadores de Lagrange o de Karush-Kuhn-Tucker. La condición (1.7) es conocida como condición de complementariedad, la condición (1.8) es conocida como condición de factibilidad dual y (1.5), (1.6) son llamadas condiciones de factibilidad primal.

Dado $\tilde{x} \in \mathbb{R}^n$ y dado $j \in \{1, \dots, m\}$, la restricción de desigualdad, $g_j(\tilde{x}) \leq 0$, se dice que es activa en el punto \tilde{x} si $g_j(\tilde{x}) = 0$; se denomina no activa si $g_j(\tilde{x}) < 0$. Se denota el conjunto de los índices de las restricciones activas por $I(\tilde{x}) = \{j : g_j(\tilde{x}) = 0\}$.

Debido a que la diferenciabilidad es un concepto local, esta permite caracterizar los mínimos locales, sin embargo, no es posible emplear esta para caracterizar los mínimos globales, considerando la condición de convexidad se obtiene que todo óptimo local también será global.

Al agregar la convexidad tanto a la función objetivo, como a las restricciones, del problema no lineal, las condiciones necesarias de optimalidad se hacen también suficientes, permitiendo caracterizar las mismas. Los siguientes resultados son prueba de ello.

Teorema 1.2.1 (Ver Cap. 3, Teo. 3.4.3, [2])

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función convexa y $S \subset \mathbb{R}^n$ con S no vacío y convexo. Considere el problema de programación convexa, $\min\{f(x) : x \in S\}$. Entonces $x^* \in S$ es una solución óptima de este problema si y sólo si, f tiene un subgradiente ξ en x^* tal que $\xi^t(x - x^*) \geq 0$, para todo $x \in S$.

Teorema 1.2.2 (Ver Cap. 4, Teo. 4.3.7, [2])

Considere el problema no lineal, $\min\{f(x) : x \in \mathbb{R}^n, h(x) = 0, g(x) \leq 0\}$, supóngase que existe una terna de vectores, $(\bar{u}, \bar{\mu}, \bar{\lambda})$, que cumple las condiciones KKT. Sea $K^+ = \{k : \bar{\lambda}_k > 0\}$ y $K^- = \{k : \bar{\lambda}_k < 0\}$. Supóngase que $f(x), g_i(x)$, para todo $i \in I(\bar{x})$ y $h_k(\bar{x})$, para todo $k \in K^+$, son funciones convexas en \mathbb{R}^n y $h_k(x)$, para todo $k \in K^-$, son funciones cóncavas en \mathbb{R}^n . Entonces \bar{x} es una solución óptima de dicho problema.

Con este resultado puede observarse que las condiciones de factibilidad y complementariedad son condiciones suficientes para la optimalidad y por ello son utilizadas en el análisis de convergencia en los diferentes métodos conocidos.

Veamos ahora otras definiciones preliminares en teoría de dualidad.

En esta teoría, se le llama problema primal al problema

$$\min_{x \in \mathbb{R}^n} \{f(x) : h(x) = 0, g(x) \leq 0\},$$

donde $x = (x_1, \dots, x_n)^t$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^\ell$ y $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Se define la función Lagrangeana como

$$\ell(x, \mu, \lambda) = f(x) + \lambda^t h(x) + \mu^t g(x). \quad (1.9)$$

Al problema primal se le puede asociar el problema dual, que consiste en maximizar la función dual, la cual viene dada por:

$$\theta(\lambda, \mu) = \inf_x \{\ell(x, \mu, \lambda)\},$$

es decir, el problema dual viene dado por

$$\text{máx}\{\theta(\lambda, \mu), \mu \geq 0, \lambda \geq 0\},$$

La teoría de dualidad es importante para la resolución de problema no lineales, ver [3]. Los resultados obtenidos dentro de esta teoría para programación convexa son útiles para desarrollar métodos que generan soluciones aproximadas a los problemas de programación no lineal.

Teorema 1.2.3 (Ver Cap. 6, Teo. 6.2.1 y Coro. 1,2,3 y 4, [2])

1. $\sup\{\theta(\lambda, \mu) : \mu \geq 0\} \leq \min\{f(x) : h(x) = 0, g(x) \leq 0\}$ (dualidad débil).
2. Si $f(x^*) \leq \theta(\lambda^*, \mu^*)$ para alguna solución factible x^* del problema primal y para alguna solución factible (λ^*, μ^*) del problema dual, entonces x^* y (λ^*, μ^*) , son respectivamente, soluciones óptimas de los problemas primal y dual.
3. Si $\sup\{\theta(\lambda^*, \mu^*) : \mu \geq 0\} = +\infty$, entonces el problema primal no tiene soluciones factibles.
4. Si $\inf\{f(x) : h(x) = 0, g(x) \leq 0\} = -\infty$, entonces $\theta(\lambda, \mu) = -\infty$ para cada $\lambda \in \mathbb{R}^\ell$ y $\mu \geq 0$.

Si un vector de multiplicadores resuelve el problema dual y la propiedad 1 del teorema anterior se satisface como igualdad, entonces las soluciones del problema Lagrangeano asociado con los multiplicadores son soluciones del problema dual. La idea básica es, por tanto, encontrar condiciones de igualdad, como es el caso en los programas convexos. El siguiente teorema es un ejemplo que permite visualizar como actúa esta propiedad de dualidad débil con solo condición de igualdad en el caso de las programación convexa.

Teorema 1.2.4 (Ver Cap. 6, Teo. 6.5.2, [2])

Supóngase que x^ y (λ^*, μ^*) son soluciones óptimas de los problemas primal y dual respectivamente y supóngase que $f(x^*) = \theta(\lambda^*, \mu^*)$. Entonces $(\lambda^*)^t g(x^*) = 0$ y x^* resuelve el problema de minimizar $\ell(x, \mu, \lambda) = f(x) + \lambda^t h(x) + \mu^+ g(x)$ sujeto a $x \in X \subset \mathbb{R}^n$.*

Capítulo 2

Métodos de Punto Proximal y de Langrangeano Aumentado.

2.1. El concepto de regularización

La idea de regularización está conectada con los problemas que tienen un mal comportamiento de la forma

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

donde f es un elemento de algún conjunto de un operador \mathbb{X} (usualmente en un espacio de funciones) y $L : \mathbb{X} \rightarrow \mathbb{X}$ es un operador.

Se quiere reemplazar $L : \mathbb{X} \rightarrow \mathbb{X}$ por otro operador regularizado $L + \lambda M$; con $\lambda \in \mathbb{R}$ y $M : \mathbb{X} \rightarrow \mathbb{X}$ donde M sea tal que el problema

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

tenga única solución f_λ para cada $\lambda > 0$, tal que f_λ se aproxime a una solución de $L(f) = 0$ cuando $\lambda \rightarrow 0$.

Este concepto se aplica en problema de optimización si nosotros tenemos $\mathbb{X} = \mathbb{R}^n$ y

$L = \nabla f$ donde f es una función convexa ($f : \mathbb{R}^n \rightarrow \mathbb{R}$) en el cual (2.1) nos queda

$$\nabla f(x) = 0 \quad (2.3)$$

ó

$$x \in \operatorname{argmin}\{f(x) : x \in \mathbb{R}^n\} \quad (2.4)$$

Este problema puede tener o no solución.

Supongase que $g : \mathbb{R}^n \rightarrow \mathbb{R}$ es 0-coerciva, es decir $\lim_{\|x\| \rightarrow \infty} g(x) = \infty$ y f es acotada inferiormente, así el límite $\|x\| \rightarrow \infty$ en (2.4) nos queda

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

tiene una única solución para cada $\lambda > 0$ ya que $f + \lambda g$ es 0-coerciva.

Esta condición de coercividad reduce el problema a hacer el estudio sobre subconjuntos compactos, teniéndose así garantizada la existencia de soluciones. Por otro lado, la convexidad estricta, implica la unicidad de tal solución. Es decir, (2.5), bajo hipótesis adecuadas, tendrá una única solución $x(\lambda)$, tal que $\lim_{\lambda \rightarrow 0^+} x(\lambda)$ existe y resuelve (2.4).

El problema con esta regularización aproximada consiste en que cuando $\lambda > 0$ tiende a 0, aunque $f + \lambda g > 0$ es estrictamente convexa y coerciva, puede que se comporte mal numéricamente, tanto como ∇f , en otras palabras, si $\nabla f(x) = 0$ es mal condicionado, entonces $(\nabla f + \lambda \nabla g) = 0$ también lo será cuando $x(\lambda)$ se aproxime a 0, aunque tenga una única solución para cada $\lambda > 0$.

Este concepto es aplicable a problemas de optimización si consideramos $\chi = \mathbb{R}^n$ y $L = \nabla f$ donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa.

Diversos métodos se han desarrollado para sobrellevar la condición planteada anteriormente, al menos teóricamente, originándose así, regularizaciones convenientes. Entre

esos se encuentran los Métodos de Punto Proximal.

2.2. Método de Punto Proximal

Considérese el algoritmo que genera una sucesión de la siguiente forma

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

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

donde λ_k es un número real que satisface $0 < \lambda \leq \tilde{\lambda}$ para $\tilde{\lambda} > 0$, $\|\cdot\|$ es una norma euclídeana.

Se probará más adelante que bajo ciertas hipótesis la sucesión generada por (2.6) y (2.7) convergen al minimizador de f .

Para ello se estudiará la definición de Fejér convergente.

Definición 2.2.1

Una sucesión $\{y^k\}$ en \mathbb{R}^n es llamado Fejér convergente a un conjunto $U \subset \mathbb{R}^n$ con respecto a la distancia euclídeana si $\|y^{k+1} - u\| \leq \|y^k - u\|$ para todo $k \geq 0$ y para todo $u \in U$.

Ejemplo 2.2.1

Sea el conjunto $U = \{(x, y) \in \mathbb{R}^2 / x = 0\}$ y $\{y^k\} = \{(\frac{1}{k}, 0)\}$, entonces, la sucesión $\{y_k\}$ es Fejér convergente en U .

Sea

$$y^k = \left(\frac{1}{k}, 0\right) \quad y \quad U = (u_1, u_2) \in U$$

Se probará que $\|y^{k+1} - u\| \leq \|y^k - u\|$ para todo $k > 0$ y $u \in U$

$$\|y^{k+1} - u\| = \sqrt{\left(\frac{1}{x+1}\right)^2 + (u_2)^2}$$

Se sabe que

$$\begin{aligned} k+1 \geq k &\Rightarrow \frac{1}{k} \geq \frac{1}{k+1} \quad \text{para todo } k \geq 0 \\ &\Rightarrow \left(\frac{1}{k+1}\right)^2 + (u_2)^2 \leq \left(\frac{1}{k}\right)^2 + (u_2)^2 \\ &\Rightarrow \sqrt{\left(\frac{1}{k+1}\right)^2 + (u_2)^2} \leq \sqrt{\left(\frac{1}{k}\right)^2 + (u_2)^2} \\ &\text{por tanto } \left\| \frac{1}{k+1} - u \right\| \leq \left\| \frac{1}{k} - u \right\| \end{aligned}$$

esto es

$$\|y^{k+1} - u\| \leq \|y^k - u\|$$

para todo $u \in U$.

Proposición 2.2.1 (Ver Prop. 2.1, [9])

Si $\{y^k\}$ es Fejér convergente en $U \neq \emptyset$ entonces $\{y^k\}$ es acotado. Si un punto de clausura y de $\{y^k\}$ pertenece a U , entonces $\lim_{k \rightarrow \infty} y^k = y$ existe.

Prueba:

Como $\{y^k\}$ es Fejér convergente a $U \neq \emptyset$ por hipótesis, esto implica que

$$\|y^{k+1} - u\| \leq \|y^k - u\| \quad \forall u \in U \quad \text{y} \quad \forall k \geq 0$$

en consecuencia, tomando $u \in U$ cualquiera, se tiene

$$\|y^k - u\| \leq \|y^{k-1} - u\| \leq \|y^{k-2} - u\| \leq \|y^{k-3} - u\| \leq \dots \leq \|y^0 - u\|$$

Así $\|y^k - u\| \leq \|y^0 - u\|$ para cualquier $u \in U$.

De lo anterior se tiene que la $\{y_k\}$ está contenida en $B(u, \|y^0 - u\|)$ y por lo tanto $\{y^k\}$ es acotada. Así $\{y^k\}$ contiene una subsucesión convergente $\{y^{j_k}\}$ tal que $\lim_{k \rightarrow \infty} y^{j_k} = y$. Además, como $\{y^k\}$ es Fejér convergente $y \in U$, en consecuencia la $\{\|y^k - y\|\}$ es decreciente, no negativa y posee una subsucesión convergente $\{\|y^{j_k} - y\|\}$ la cual converge a cero, por lo tanto $\{\|y^k - y\|\}$ también converge a 0, es decir

$$\lim_{k \rightarrow \infty} \|y^k - y\| = 0 \quad \text{implica} \quad \lim_{k \rightarrow \infty} y^k = y$$

■

Teorema 2.2.1 (ver Teo. 2.1, [9])

Sea $f : \mathbb{R} \rightarrow \mathbb{R}$ convexa y continuamente diferenciable. Supongamos que $U \neq \emptyset$ es el conjunto de minimizadores de f sobre \mathbb{R}^n . Entonces la sucesión $\{x^k\}$ generada por (2.6), (2.7) converge a un punto $x^* \in U$.

Prueba:

Se realizará la prueba en 4 pasos:

Paso 1:

Se probará que la sucesión está bien definida

$$x^0 \in \mathbb{R}^n$$

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

Procediendo por inducción

Para $k = 1$ se cumple, ya que como $x^0 \in \mathbb{R}^n$ se puede asegurar la existencia de

$$x^1 = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \{f(x) + \lambda_1 \|x - x^0\|\}$$

Supongase que se cumple para k , y se probará para $k + 1$ es decir,

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

Además $f_k = f(x) + \lambda_k \|x - x^k\|^2$. Verificando que x^{k+1} existe y es único.

Como $U \neq \emptyset$, f alcanza un mínimo. Así este está acotado inferiormente, y además $\|x - x^k\|^2$ es estrictamente convexa por lo que

$$\lim_{k \rightarrow \infty} f(x) + \lambda_k \|x - x^k\|^2 = \infty \quad \lim_{k \rightarrow \infty} f_k(x) = \infty$$

Observe que si f_k es continua y el minimizador está en (2.7), reduciendo el conjunto a un conjunto compacto, f_k logre el mínimo.

f es convexa y $\lambda_k \|x - x^k\|^2$ es estrictamente convexa, entonces f_k es estrictamente convexa, así esto tiene un mínimo, es decir x^{k+1} es único.

Paso 2:

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 \quad \text{para todo } k \geq 0 \quad \text{y} \quad \text{para todo } \bar{x} \in U$$

En efecto; para $\bar{x} \in U$, cualquiera

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

Por otro lado x^{k+1} es un minimizador se tiene que

$$0 = \nabla f_k(x^{k+1}) = \nabla f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k) \quad (2.9)$$

Observar que

$$f_k = f(x) + \lambda_k \|x - x^k\|^2$$

Donde

$$\nabla f(x^{k+1}) = \nabla f(x^{k+1}) + \nabla(\lambda_k \|x - x^k\|^2)$$

donde

$$\begin{aligned}\lambda_k \|x - x^k\|^2 &= \lambda_k \langle x^{k+1} - x^k, x^{k+1} - x^k \rangle \\ &= \lambda_k \sum_{i=1}^n (x^{i+1} - x^i)^2 \\ \lambda_k \nabla \sum_{i=1}^n (x^{i+1} - x^i)^2 &= 2\lambda_k (x^{k+1} - x^k)\end{aligned}$$

Luego, de (2.9) se obtiene

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

Sustituyendo (2.8) en (2.10) se tiene

$$\|x^k - \bar{x}\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - \bar{x}\|^2 = \frac{1}{\lambda_k} \langle \nabla f(x^{k+1}), x^{k+1} - \bar{x} \rangle \quad (2.11)$$

Como f es convexa y diferenciable, se sabe por definición que

$$\begin{aligned}f(x^{k+1}) &\leq f(\bar{x}) + \langle \nabla f(x^{k+1}), x^{k+1} - \bar{x} \rangle \\ 0 &\leq \frac{1}{\lambda_k} [f(x^{k+1}) - f(\bar{x})] \leq \frac{1}{\lambda_k} \langle \nabla f(x^{k+1}), x^{k+1} - \bar{x} \rangle\end{aligned}$$

ya que $f(\bar{x}) \leq f(x^{k+1})$ para todo k

De aquí

$$\|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 - \|x^{k+1} - \bar{x}\|^2 \geq 0$$

Paso 3: $\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0$ en efecto;

Del paso 2 se tiene que

$$0 \leq \|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 \quad \text{para todo } k \geq 0 \quad \text{para todo } \bar{x} \in U$$

Observar que la sucesión $\{\|x^k - \bar{x}\|\}$ es decreciente y no-negativa, por lo tanto convergente, además

$$\|x^k - \bar{x}\| \longrightarrow 0 \quad \text{y} \quad \|x^{k+1} - \bar{x}\| \longrightarrow 0$$

por tanto

$$\|x^{k+1} - x^k\| \longrightarrow 0$$

Paso 4: Se probará que cualquier punto clausura de $\{x^k\}$ está en U .

La existencia de punto clausura se sigue del paso 2 (fejer convergencia) y la **proposición 2.1**

Sea \bar{x} un punto de clausura de $\{x^k\}$, entonces existe una subsucesión $\{x^{j_k}\}$ de $\{x^k\}$ tal que $\lim_{k \rightarrow \infty} x^{j_k} = \bar{x}$

Luego, de (2.2) se tiene que

$$\nabla f(x^{j_k+1}) = 2\lambda_{j_k}(x^{j_k} - x^{j_k+1}) \quad (2.12)$$

Del paso 3 se obtiene que

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

tomando el límite en (2.12) se obtiene que

$$\nabla f(\bar{x}) = 0$$

Por la convexidad de f , $\bar{x} \in U$, y los pasos 2 y 4, se obtiene que satisfacen las hipótesis de la **Proposición 2.1**, por lo que existe $x^* \in U$ tal que $x^* = \lim_{k \rightarrow \infty} x^k$. ■

Este teorema es la versión clásica de la convergencia del Método de Punto Proximal. Inicialmente, en 1965, Moreau [20], ofrece esta regularización, concretamente propone la siguiente función:

$$F(y) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2} \|x - y\|^2 \right\},$$

que luego es modificada por Yosida, agregando el parámetro λ convirtiéndose en la regularización Moreau-Yosida, conocida por la expresión siguiente:

$$F_\lambda(y) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2} \lambda \|x - y\|^2 \right\}$$

Obsérvese que el Método de Punto Proximal Clásico sustituye el parámetro λ por una sucesión $\{\lambda^k\}$ acotada y de números positivos. Martinet [21] en 1970 introduce este método en el estudio de la programación convexa y posteriormente Rockfellar [22] generaliza la técnica para operadores monótonos maximales.

2.3. Operadores Monótonos Maximales

Definición 2.3.1

Un operador $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es llamado monótono si y sólo si $\langle x - y, T(x) - T(y) \rangle \geq 0$ para cada $x, y \in \mathbb{R}^n$.

Un ejemplo es $T = \nabla f$ con $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable y convexo. Es claro que el gradiente sería sustituido por subgradiente en el caso en que la función no sea diferenciable.

Definición 2.3.2

Un operador $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es llamado monótono maximal si y sólo si T es monótono y para todo T' monótono tal que $T(x) \subset T'(x)$ para cada x , se tiene que $T = T'$.

El problema de interés es encontrar el cero del operador monótono maximal si

$T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ entonces x es cero de T sii $0 \in T(x)$. Si $T : \mathbb{R}^n \rightarrow \mathbb{P}(\mathbb{R}^n)$, x es cero de T sii $0 \in T(x)$.

Para f convexa y $T = \partial f$ se tiene que x es cero de T sii x es un minimizador de f , ya que $0 \in T(x)$ se tiene para todo y que:

$$0 = \langle 0, y - x \rangle \leq f(y) - f(x)$$

así $f(x) \leq f(y)$ para todo $y \in \mathbb{R}^n$, obteniéndose que el problema de encontrar ceros de el operador monótono maximal generaliza el problema de minimización de funciones convexas.

En el caso del Método de Punto Proximal Clásico, para que $\{x^{k+1}\}$ resuelva el (2.2), haciendo $f_k(x) = f(x) + \lambda_k \|x - x^k\|^2$, se cumple:

$$0 = \nabla f_k(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k)$$

o bien,

$$\nabla f(x^{k+1}) = 2\lambda_k(x^k - x^{k+1}) \quad (2.13)$$

Lo que sugiere una extensión natural a operadores monótonos, dada por

$$\lambda_k(x^k - x^{k+1}) \in T(x^{k+1}) \quad (2.14)$$

lo cual equivale a

$$x^{k+1} \in (I + \frac{1}{\lambda_k} T)^{-1}(x^k) \quad (2.15)$$

Antes de establecer el resultado que muestra la convergencia del Método de Punto Proximal para operadores monótonos, estableceremos algunas definiciones y teoremas previos.

Definición 2.3.3

Un operador $T : \mathbb{R}^n \rightarrow P(\mathbb{R}^n)$ es llamado *no-expansivo* si

$$\|T(x) - T(y)\| \leq \|x - y\|$$

para todo $x, y \in \mathbb{R}^n$.

Definición 2.3.4

Sea $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es firmemente no-expansivo sii

$$\|P(x) - P(y)\|^2 \leq \|x - y\|^2 - \|(x - y) - (P(x) - P(y))\|^2$$

Proposición 2.3.1 (Ver Prop. 4.1, [9])

Si P es firmemente no-expansivo, entonces la sucesión definida por $x^0 \in \mathbb{R}^n$, $x^{k+1} = P(x^k)$ es Fejér convergente al conjunto de puntos fijos de P :

Prueba:

P es firmemente no-expansivo si

$$\|P(x) - P(y)\|^2 \leq \|x - y\|^2 - \|(x - y) - (P(x) - P(y))\|^2$$

Aplicaremos esta definición con $y = \bar{x}$ tal que $P(\bar{x}) = \bar{x}$, $x = x^k$, $P(x) = x^{k+1}$. Se obtiene

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

Note que esta es la propiedad tres puntos usada en el paso 2 en la prueba del **Teorema 2.2.1**.

■

Definición 2.3.5

Sea $P : \mathbb{R}^n \rightarrow \mathbb{P}(\mathbb{R}^n)$ diremos que P es sobre si para todo $y \in \mathbb{R}^n$ existe, $x \in \mathbb{R}^n$ tal que $y \in P(x)$ y P es 1-1. Si para todo $x \neq y$ se cumple $P(x) \cap P(y) = \emptyset$.

Teorema 2.3.1 (Ver Teo. 4.1, [9] Teorema de Minty)

Si $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es monótono y maximal y $\mu > 0$ entonces $I + \mu T$ es uno a uno y sobre, además $(I + \mu T)^{-1}$ es firmemente no-expansivo.

Lema 2.3.1 (Ver Lema 4.1, [9])

Si $\lim_{k \rightarrow \infty} y^k = \bar{y}$, $\lim_{k \rightarrow \infty} z^k = \bar{z}$, T es monótono maximal y $y^k \in T(z^k)$, entonces $\bar{y} \in T(\bar{z})$.

Teorema 2.3.2

Si $T : \mathbb{R}^n \rightarrow \mathbb{P}(\mathbb{R}^n)$ es maximal monótono y existe \bar{x} tal que $0 \in T(\bar{x})$, entonces la sucesión $\{x^k\}$ definida por

$$\begin{aligned} x^0 &\in \mathbb{R}^n \\ x^{k+1} &\in \left(I + \frac{1}{\lambda_k} T \right)^{-1} (x^k) \end{aligned}$$

con $0 < \lambda_k < \tilde{\lambda}$ converge a un vector x^* tal que $0 \in T(x^*)$.

Prueba:

Paso 1:

En efecto: $\{x^k\}$ está bien definida, lo cual es inmediato por el teorema de Minty.

En este caso utilice el hecho de que $I + \mu T$ es sobre con lo que se garantiza la existencia de $x^{k+1} \in (I + \mu T)^{-1}(x^k)$.

Paso 2:

En efecto; $\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2$ para todo k y \bar{x} tal que $0 \in T(\bar{x})$.

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

de donde

$$\|x^k - \bar{x}\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - \bar{x}\|^2 = 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x} \rangle$$

Se sabe que

$$\lambda_k(x^k - x^{k+1}) \in T(x^{k+1})$$

Además, por hipótesis $0 \in T(\bar{x})$.

Por tanto

$$\|x^k - \bar{x}\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - \bar{x}\|^2 = \frac{2}{\lambda_k} \langle \lambda_k(x^k - x^{k+1}) - 0, x^{k+1} - \bar{x} \rangle$$

para todo x^{k+1}, \bar{x} y $\lambda_k(x^k - x^{k+1}) \in T(x^{k+1}), 0 \in T(\bar{x})$

Así se cumple la propiedad tres puntos y además la fejer convergencia a $T(\bar{x})$

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

.

Paso 3:

En efecto; $\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0$

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

La sucesión $\{\|x^k - \bar{x}\|^2\}$ es decreciente y no negativa, por lo tanto convergente además,

$$\|x^k - \bar{x}\| \rightarrow 0 \quad \text{y} \quad \|x^{k+1} - \bar{x}\| \rightarrow 0$$

por lo que

$$\|x^{k+1} - x^k\| \rightarrow 0$$

Paso 4:

En efecto; $\{x^k\}$ tiene puntos de acumulación, entonces estos son ceros de T .

$\{x^k\}$ tiene punto de acumulación \hat{x} por ser fejer convergente, por tanto existe una sub-sucesión $\{x^{j_k}\}$ de $\{x^k\}$ tal que $\lim_{k \rightarrow \infty} x^{j_k} = \hat{x}$, luego, por el paso 3 $\lim_{k \rightarrow \infty} x^{j_{k+1}} = \hat{x}$, pero

$$\lambda_k(x^{j_k} - x^{j_{k+1}}) \in T(x^{j_{k+1}})$$

es claro que $\lambda_{j_k}(x^{j_k} - x^{j_{k+1}}) \rightarrow 0$ porque $\lambda_k < \tilde{\lambda}$, pero no se puede tomar límite en $T(x^{j_{k+1}})$.

Usando

$$y^k = \lambda_{j_k}(x^{j_{k+1}}), \bar{y} = 0 \quad z^k = x^{j_{k+1}}\bar{z} = \hat{x}$$

Por el **Lema (2.3.1)** se tiene que $0 \in T(\hat{x})$; esto prueba el paso 4.

Por lo que el paso 2 y 4 conjuntamente con la fejer convergencia demuestran el teorema. ■

2.4. Lagrangeano Aumentado.

Los métodos de multiplicadores transforman un problema de minimización con restricciones, en una secuencia de problemas de minimización sin restricciones. Dicha transformación se realiza adicionando a la función objetivo del problema original una función de penalidad la cual involucra a un parámetro $\mu \in \mathbb{R}^n$.

Existen diversas versiones de estos métodos. En este trabajo se consideran dos tipos, a saber el método de multiplicadores clásicos presentado por Rockafellar en [16] y el método de langrangeano aumentado considerado en Iusem en [12].

Considerar el problema primal de minimización

$$f^* = \inf \{ f(x) : g_i(x) \leq 0 \quad i = 1, \dots, m \} \quad (2.16)$$

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

El Lagrangeano clásico para este problema es $L : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ definido por:

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

El método de Lagrangeano clásico consiste en generar dos sucesiones, $\{x^k\} \subset \mathbb{R}^n$ y $\{\mu^k\} \subset \mathbb{R}^m$, de la siguiente forma; dado $\mu^k \geq 0$, se genera $x^k = \operatorname{argmin} \left\{ L(x, \mu^k), \text{ con } x \in \mathbb{R}^n \right\}$, actualizando μ^k de alguna manera conveniente.

Computacionalmente los problemas están relacionados con la discontinuidad de $L(x, \cdot)$, surgiendo así la necesidad de mejorarlos.

Se estudiará ahora el método de Lagrangeano Aumentado, considerado por Iusem en [12].

Para $x \in \mathbb{R}^n$ defínase x^+ de la siguiente forma $x_j^+ = \max\{x_j, 0\}$. Tomese $\alpha > 0$ y el Lagrangeano aumentado clásico

$$L_\alpha(x, \mu) = f(x) + \frac{1}{\alpha} \sum_{i=1}^m \{[(\mu_i + 2\alpha g_i(x))^+]^2 - \mu_i^2\}$$

$L_\alpha(x, \mu)$ es diferenciable si y sólo si f y g_i lo son, de hecho:

$$\nabla_1 L_\alpha(x, \mu) = \nabla f(x) + \sum_{i=1}^m [\mu_i + 2\alpha g_i(x)]^+ \nabla g_i(x)$$

El método en consideración genera dos sucesiones $\{x^k\} \subset \mathbb{R}^n$ y $\{\mu^k\} \subset \mathbb{R}^m$ de la siguiente forma:

Dados

$$\mu^0 \in \mathbb{R}_+^m \tag{2.18}$$

$$x^k = \operatorname{argmin} L_\alpha(x, \mu^k), x \in \mathbb{R}^n \tag{2.19}$$

$$\mu_i^{k+1} = [\mu_i^k + 2\alpha g_i(x^k)]^+ \tag{2.20}$$

La convergencia de la sucesión $\{x^k\}$ no puede ser obtenida inmediatamente de la teoría de punto proximal. De hecho, se requieren hipótesis adicionales, como por ejemplo, además de otros, la condición de Slater; que exista x tal que $g_i(\bar{x}) < 0$ para todo $i = 1, \dots, m$

y como $\{x^k\}$ minimiza el Lagrangeano $L(., \mu^{k+1})$, entonces la sucesión $\{x^k\}$ converge al minimizador x^* y así, el par (x^*, μ^*) es un punto silla para L resultando por la teoría de dualidad y convexidad que x^* es la solución de (2.16).

2.5. Penalización

Considere el problema

$$f^* = \inf\{f(x) : x \in \bar{S} \subset \mathbb{R}^n\} \quad (2.21)$$

con $\bar{S} \subset \mathbb{R}^n$.

Para resolver (2.21) se puede considerar una función penalidad g tal que $g(x) = +\infty$, si x no pertenece \bar{S} .

Para $r > 0$, el problema regularizado:

$$\min_{x \in \mathbb{R}^n} \{f(x) + rg(x)\} \quad (2.22)$$

Téndra su solución en \bar{S} (si (2.21) tiene solución). Si (2.21) tiene solución y g es escogido adecuadamente, entonces (2.22) tendrá una única solución $x(r)$ y bajo hipótesis adicionales es posible probar que $\lim_{r>0} x(r)$ y resuelve (2.21). Como en la regularización, aquí también puede ocurrir que el problema (2.22) este mal comportado para valores pequeños. Es decir, cuando $x \in \delta\bar{S}$ y $g(x)$ es muy grande, $rg(x)$ puede ser que se comporte inadecuadamente cuando $r \rightarrow 0$. Lo que se requiere es combinar la idea de regularización con la de punto proximal para conseguir convergencia cuando r es cercano a 0.

En concreto, llamamos penalización exacta a la escogencia de una función g tal que $x(\bar{r})$, obtenida para (2.22) resuelva para (2.21) para algún \bar{r} fijo aunque la existencia teórica de tal \bar{r} sea posible, el valor de \bar{r} no tiene que ser conocido, sin embargo el mal

comportamiento comentado arriba pudiera permanecer.

Por lo anterior expuesto el método de punto proximal, en este sentido, trabaja con un r arbitrario o una sucesión $\{r_k\}$ acotada por debajo para evitar dichos problemas.

2.6. Distancia de Bregman y Método de Punto Proximal

Definición 2.6.1

Sea S un subconjunto no vacío de \mathbb{R}^n . Una función $D : S \times S \rightarrow \mathbb{R}_+$, con $S = \text{int}\bar{S}$ tal que: $D(x, y) = 0$ si y solo si $x = y$ y $D(x, y) \geq 0$ para cada $x, y \in S$ es llamada una casi-distancia en S .

Considérese dos clases de tales distancias: Las Distancia de Bregman y φ -divergencia. En ambos casos el método de punto proximal consiste en generar una sucesión $\{x^k\}$ con $x^0 \in S$ y $x^{k+1} = \underset{x \in \mathbb{R}^n}{\text{argmin}} \{f(x) + r_k D(x, x^k)\}$.

Definición 2.6.2

Sea S un subconjunto de \mathbb{R}^n abierto y convexo y \bar{S} su clausura.

Considere una función real convexa h definida sobre \bar{S} y sea $\bar{D}_h : S \times S \rightarrow \mathbb{R}$ dada por

$$D_h(x, y) = h(x) - h(y) - \nabla h(y)^T(x - y)$$

h es llamada función de Bregman y D_h la distancia de Bregman inducida por h si satisfacen las siguientes condiciones:

B_1) h es continuamente diferenciable sobre S

B_2) h es estrictamente convexa y continua sobre \bar{S}

$B_3)$ Para todo $\delta \in \mathbb{R}$ los conjuntos parciales de sub-nivel.

$$\Gamma_1(y, \delta) = \{x \in \bar{S} : D_h(x, y) \leq \delta\}$$

$$\Gamma_2(x, \delta) = \{y \in S : D_h(x, y) \leq \delta\}$$

son acotados para todo $y \in S$, todo $x \in \bar{S}$, respectivamente.

$B_4)$ 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

$$\lim_{k \rightarrow \infty} D_h(x^k, y^k) = 0 \text{ entonces } \lim_{k \rightarrow \infty} x^k = y^*$$

S es llamada la zona de h .

Es fácil ver que $D(x, y) \geq 0$ para todo $x \in \bar{S}$ y que $D(x, y) = 0$ si $x = y$ esta se obtiene del hecho que h es estrictamente convexa, esto es

$$f(x) \langle f(y) + \nabla h(x)(x - y) \rangle \text{ siempre que } x \neq y$$

por lo tanto,

$$D(x, y) \geq 0 \text{ para } x \in \bar{S} \text{ y } D(x, y) = 0 \text{ sii } x = y$$

Definición 2.6.3

Se dice que una función h de Bregman es coerciva acotada en la frontera si para

$$\{y^k\} \subset S$$

tal que

$$\lim_{k \rightarrow +\infty} y^k = y \in \delta S$$

entonces

$$\sum_{k \rightarrow \infty} (\nabla h(y^k))^t (x - y^k) = -\infty \text{ para todo } x \in S.$$

h es llamado zona coerciva si para toda $y \in \mathbb{R}^n$ existe $x \in S$ tal que $\nabla h(x) = y$.

Ejemplo 2.6.1

$S = \mathbb{R}_{++}^n$, $h(x) = \sum_{j=1}^n x_j \log x_j$; continua sobre $\delta\mathbb{R}_+^n$ con la convención $0 \log 0 = 0$.

Se tiene que

$$D_h(x, y) = \sum_{j=1}^n \left(x_j \log \left(\frac{x_j}{y_j} \right) - x_j + y_j \right)$$

Ejemplo 2.6.2

$S = \mathbb{R}^h$, $h(x) = x^T M x$, con $M \in \mathbb{R}^{h \times h}$ simétrica y definida positiva, halle $D_h(x, y)$.

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle$$

Así el $\nabla h(y)$ donde $h(y) = y^T M y$

$$\begin{aligned} y^T M y = \langle M y, y \rangle &= \left\langle \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \dots & \dots & \dots & \dots \\ m_{n1} & \dots & \dots & m_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle \\ &= \left\langle \begin{pmatrix} m_{11}y_1 + m_{12}y_2 + \dots + m_{1n}y_n \\ m_{21}y_1 + m_{22}y_2 + \dots + m_{2n}y_n \\ \vdots \\ m_{n1}y_1 + m_{n2}y_2 + \dots + m_{nn}y_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle \\ &= \left\langle \begin{pmatrix} \sum_{j=1}^n m_{1j}y_j \\ \vdots \\ \sum_{k=1}^n m_{nk}y_k \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{j=1}^n m_{1j}y_j y_1 + \dots + \sum_{j=1}^n m_{nj}y_j y_n = \sum_{k=1}^n m_{kj}y_j y_k \end{aligned}$$

$$\nabla h(y) = 2y^T M$$

$$\begin{aligned} D_n(x, y) &= x^T Mx - y^T My - \langle (2y^T M)^T, (x - y) \rangle \\ &= \langle Mx, y \rangle - \langle My, y \rangle - \langle 2My, x - y \rangle \\ &= \langle M(x + y), x - y \rangle - \langle 2My, x - y \rangle \\ &= \langle M(x + y - 2y), x - y \rangle \\ &= \langle M(x - y), x - y \rangle = (x - y)^T M(x - y) \end{aligned}$$

Proposición 2.6.1 (ver [9], Prop. 9.1)

Si h es una función de Bregman con zona S , entonces

$$i) D_h(x, y) - D_h(x, z) - D_h(z, y) = \langle \nabla h(y) - \nabla h(z), z - x \rangle$$

$$ii) \nabla_x D_h(x, y) = \nabla h(x) - \nabla h(y) \text{ para todo } x, y \in S$$

$$iii) D_h(\cdot, y) \text{ es estrictamente convexo para todo } y \in S$$

Demostración:

i) De la definición de distancias de Bregman se tiene que

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h^T(y), x - y \rangle$$

$$D_h(x, z) = h(x) - h(z) - \langle \nabla h^T(z), x - z \rangle$$

$$D_h(z, y) = h(z) - h(y) - \langle \nabla h^T(y), z - y \rangle$$

Así

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

$$\text{ii) } \nabla_x D_h(x, y) = \nabla_x(h(x) - h(y)) - \nabla_x(\nabla h(y)(x - y))$$

$$\nabla_x h(x) - \nabla_x \sum \nabla h(y_i)x_i = \nabla h(x) - \nabla h(y) \quad x, y \in S$$

iii) $D_h(\cdot, y)$ es estrictamente convexo para todo $y \in S$

Se sabe que f es convexa y diferenciable si

$$D_h(x, z) > D_h(y, z) + \langle \nabla D_h(y, z), x - y \rangle$$

h es estrictamente convexa

$$h(x) > h(y) + \langle \nabla h(y), x - y \rangle$$

$$h(x) - h(z) - \langle \nabla h(z), x - z \rangle > h(y) - h(z) - \langle \nabla h(z), x - z \rangle + \langle \nabla h(y), x - y \rangle$$

$$D_h(x, z) > h(y) - h(z) - \langle \nabla h(z), y - z \rangle + \langle \nabla h(z), y - z \rangle - \langle \nabla h(z), x - z \rangle + \langle \nabla h(y), x - y \rangle$$

$$D_h(x, z) > D_h(y, z) - \langle \nabla h(z), x - y \rangle + \langle \nabla h(y), x - y \rangle$$

$$D_h(x, z) > D_h(y, z) + \langle \nabla h(y) - \nabla h(z), x - y \rangle$$

$$D_h(x, z) > D_h(y, z) + \langle \nabla D_h(y, z), x - y \rangle \quad \blacksquare$$

La Distancia de Bregman son usadas para generar un método de punto proximal.

El problema de interés es

$$f^* = \inf\{f(x) : x \in \bar{S} \subset \mathbb{R}^n\} \quad (2.23)$$

con $S \subset \mathbb{R}^n$ abierto y convexo, \bar{S} la clausura de S y f continua y convexa sobre \bar{S} .

El método de punto proximal con distancia de Bregman S definida como

$$x^0 \in S \quad (2.24)$$

$$x^{k+1} = \operatorname{argmim}\{f(x) + \lambda_k D_h(x, x^k)\} \quad (2.25)$$

donde h es una función de Bregman con zona S y λ_k satisface

$$0 \leq \lambda_k \leq \tilde{\lambda} \quad (2.26)$$

para algún $\tilde{\lambda} > 0$.

El siguiente teorema establece la convergencia de la sucesión generada por (2.24)-(2.25)

Teorema 2.6.1 (ver [9], Teo. 10.1)

Si el problema (2.23) tiene solución y h es coerciva acotada, con respecto a S , entonces la sucesión $\{x^k\}$ generada por (2.24) y (2.25) converge a la solución x^ de problema (2.23).*

Demostración:

Paso 1:

Se probará en primer lugar que la sucesión generada por (2.25) está bien definida y contenida en S .

En efecto, como f es continua en \bar{S} , es acotada, por lo que se puede considerar un β como una cota inferior de f , por lo que

$$f_x(x) = f(x) + \lambda_k D_h(x, x^k) \geq \beta + \lambda_k D_h(x, x^k) \quad \text{para todo } x \in \bar{S}$$

por la propiedad (B_3) se sigue que los conjuntos de nivel de asociados a D_h son acotados, por lo que $f_k(x)$ es acotada, y por lo tanto la minimización de (2.25) se reduce a un conjunto compacto, en consecuencia el mínimo es obtenido.

Por otro lado, como f es convexa, h es estrictamente convexa. Por **Proposición 2.5.1(iii)** se tiene que $D_h(\cdot, y)$ es estrictamente convexa para todo $y \in S$, y así f_k es estrictamente convexa, por tanto x^{k+1} es un minimizador único de f_k .

Véase ahora que $x^{k+1} \in S$

Como x^{k+1} es único minimizador de f_k , se deduce que x^{k+1} es el único $x \in \bar{S}$ que verifica

$$\begin{aligned} 0 &\in \partial(f(x^{k+1}) + \lambda_k D_h(x^{k+1}, x^k)) \\ 0 &\in (\partial(f(x^{k+1}) + \lambda_k \partial(D_h(x^{k+1}, x^k))) \end{aligned}$$

Luego, como D_h es diferenciable

$$\begin{aligned} 0 &\in \partial f(x^{k+1}) + \lambda_k \nabla_x(D_h(x^{k+1}, x^k)) \\ 0 &\in \partial f(x^{k+1}) + \lambda_k (\nabla h(x^{k+1}) - \nabla h(x^k)) \\ 0 &\in \partial f(x^{k+1}) + \lambda_k \nabla h(x^{k+1}) - \lambda_k \nabla h(x^k) \end{aligned}$$

Así

$$\lambda_k \nabla h(x^k) \in \partial(f + \lambda_k h)(x) \quad (2.27)$$

Se probará, bajo (B_5) , $\partial(f + \lambda_k h)(x) = \emptyset$ para todo $x \in \partial S$ en consecuencia con (2.5) se tiene $x^{k+1} \in S$.

Tomando $x \in \partial S$, supongamos que existe $\epsilon \in \partial(f + \lambda_k h)(x)$, $z \in \bar{S}$ y definase

$$y^\ell = (1 - \epsilon_\ell)x + \epsilon_\ell z; \quad (2.28)$$

donde $\lim_{\ell \rightarrow \infty} \epsilon_\ell = 0$, entonces $y^\ell \in X$, y X es convexo.

Además

$$\lim_{\ell \rightarrow \infty} y^\ell = x$$

Por (2.28) se obtiene:

$$c_\ell \xi^\ell(z - x) = \xi^\ell(y^\ell - x) \quad (2.29)$$

Como $\xi \in \partial(f + h/c_k)$, se deduce:

$$\xi^\ell(y^\ell - x) \geq f(y^\ell) - f(x) + \frac{(h(y^\ell) - h(x))}{c_k} \quad (2.30)$$

Como h es estrictamente convexa y continuamente diferenciable, por la propiedades (1),(2). Se deduce:

$$f(y^\ell) - f(x) + \frac{(h(y^\ell) - h(x))}{c_k} \geq f(y^\ell) - f(x) + \nabla \frac{h(y^\ell)^t(y^\ell - x)}{c_k} \quad (2.31)$$

usando nuevamente (2.28) y la convexidad de f se obtiene:

$$\frac{f(y^\ell) - f(x) + \nabla h(y^\ell)^t(y^\ell - x)}{c_k} \geq \epsilon_\ell(f(z) - f(x)) + \frac{\epsilon_\ell}{c_k(1 - \epsilon_\ell)} \nabla h(y^\ell)^t(z - y^\ell) \quad (2.32)$$

Por lo cual de (2.29), (2.30), (2.31) y (2.32) se tiene que:

$$c_\ell \xi^t(z - x) \leq \epsilon_\ell(f(z) - f(x)) + \frac{\epsilon_\ell}{c_k(1 - \epsilon_\ell)} \nabla h(y^\ell)^t(z - y^\ell) \quad (2.33)$$

En consecuencia:

$$c_\ell(1 - \epsilon_\ell)[f(x) - f(z) + \xi^t(z - x)] \leq \nabla h(y^\ell)^t(z - y^\ell) \quad (2.34)$$

Como $\lim_{\ell \rightarrow \infty} y^\ell = x$ y además x está en la frontera de S , asimismo h es coerciva en la frontera, por lo cual el lado derecho (2.34) tiende a $-\infty$ cuando $\ell \rightarrow \infty$, mientras el lado izquierdo tiene un límite finito, lo cual es contradictorio, por lo tanto $\partial(f + D_h) = \emptyset$, para todo x en la frontera de X , en consecuencia $x^{k+1} \in S$

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 toda solución \bar{x} de (2.2.3).

En efecto;

$$D_h(x, y) - D_h(x, z) - D_h(z, y) = \langle \nabla h(y) \rangle - \nabla h(z), z - x$$

con

$$x = \bar{x}, y = x^k, z = x^{k+1}$$

obteniendo que

$$D_h(\bar{x} - x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \langle \nabla h(x^k) \rangle - \nabla h(x^{k+1}), x^{k+1} - \bar{x} \quad (2.35)$$

De (2.3.5) se tiene que

$$\begin{aligned}
0 &\in \partial(f + \lambda_k D(\cdot, x^k))(x^{k+1}) \\
0 &\in \partial f(x^{k+1}) + \lambda_k \nabla D_h(x^{k+1}, x^k) \\
-\lambda_k [\nabla h(x^{k+1}) - \nabla h(x^k)] &\in \partial f(x^{k+1}) \\
\lambda_k [\nabla h(x^k) - \nabla h(x^{k+1})] &\in \partial f(x^{k+1})
\end{aligned} \tag{2.36}$$

Sea $y^k = \lambda_k (\nabla h(x^k) - \nabla h(x^{k+1}))$

De (2.36)

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

Luego, por definición de subgradiente

$$\begin{aligned}
f(\bar{x}) &\geq (x^{k+1}) + (y^k, \bar{x} - x^{k+1}) \Rightarrow f(\bar{x}) - f(x^{k+1}) \geq (y^k, \bar{x} - x^{k+1}) \\
\Rightarrow \frac{1}{\lambda_k} (f(x^{k+1}) - f(\bar{x})) &\leq \frac{1}{\lambda_k} (y^k, x^{k+1} - \bar{x})
\end{aligned}$$

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

De aquí

$$-D_h(x^{k+1}, x^k) D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) \geq \frac{1}{\lambda_k} (f(x^{k+1}) - f(\bar{x})).$$

Luego, como \bar{x} es mínimo se tiene que

$$\begin{aligned}
D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) &\geq 0 \\
\text{por tanto } D_h(\bar{x}, x^{k+1}) &\leq D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k)
\end{aligned} \tag{2.37}$$

Paso 3:

$\{x^k\}$ es acotada y $\lim_{k \rightarrow \infty} x^{j^k} = \hat{x}$ implica que $\lim_{k \rightarrow \infty} x^{k+1}$.

En efecto;

Por el paso 2 se tiene que $\{D_h(\bar{x}, x^k)\}$ es decreciente y no negativa y por lo tanto convergente.

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

Como $\{D_h(\bar{x}, x^k)\}$ es decreciente, se tiene que $D_h(\bar{x}, x^k) \leq D_h(x, x^0)$.

Por lo tanto $\{x^k\}$ es acotado, en virtud que h es acotada (B_3).

Si $\lim_{k \rightarrow \infty} x^{j_k} = \hat{x}$ una subsucesión de $\{x^k\}$, entonces $\lim_{k \rightarrow \infty} x^{j_{k+1}} = \hat{x}$ por (B_5).

Paso 4:

Se probará que los puntos de acumulación de $\{x^k\}$ son solución del problema (2.36).

Sea una solución de (2.25), \bar{x} y \hat{x} un punto clausura de $\{x^k\}$ luego, como esta es acotada, existe una subsucesión $\{x^{j_k}\}$ tal que

$$\lim_{k \rightarrow \infty} x^{j_k} = \hat{x},$$

en el paso 3 se probó que $\lim_{k \rightarrow \infty} x^{j_{k+1}} = \hat{x}$.

$$0 \leq \frac{1}{\lambda_k} (f(x^{j_{k+1}}) - f(\bar{x})) \leq D_h(\bar{x} - x^k) - D_h(x - x^{j_{k+1}}) - D_h(x^{j_{k+1}} - x^k).$$

De (2.36) y la desigualdad anterior se tiene que

$$0 \leq \frac{1}{\lambda_k} (f(x^{j_{k+1}}) - f(\bar{x})) \leq \frac{1}{\lambda_k} (f(x^{j_{k+1}}) - f(\bar{x})) \leq D_h(\bar{x}, x^{j_k}) - D_h(\bar{x}, x^{j_{k+1}}) - D_h(x^{j_{k+1}}, x^{j_k})$$

Como $D_h(\bar{x} - x^{j_k}) - D_h(\bar{x}, x^{j_{k+1}}) - D_h(x^{j_{k+1}}, x^{j_k}) \rightarrow 0$ cuando $k \rightarrow \infty$ obteniendo que $\lim_{k \rightarrow \infty} f(x^{j_{k+1}}) = f(\hat{x})$, además, por (β_5) se obtiene que $f(\hat{x}) = f(\bar{x})$. ■

Un ejemplo que se debe resaltar es el (2.6.1), el cual es de la distancia entrópica o divergencia Kullback-Leibler.

Otro ejemplo que se obtiene al considerar el función de Bregman $h(x) = \frac{1}{2}\|x\|^2$ resultando que $D_h(x, y) = \frac{1}{2}\|x - y\|^2$, recuperándose el núcleo cuadrático del método de punto proximal clásico.

2.7. φ -Divergencia y Método de Punto Proximal

Continuando con el debate y aplicaciones del método proximal, se señalara el trabajo de Teboulle [11], estudiando las propiedades básicas de la función Entrópica proximal, la cual está basada sobre la imitación de la función proximal de Moreau, donde la distancia cuadrática del método proximal clásico es reemplazada por la φ -divergencia de Csiszár (1967).

Estudiando así otra clase de distancias, que son denotadas por $d_\varphi(\cdot, \cdot)$ y están definidas sobre el ortante positivo de \mathbb{R}^n .

Definición 2.7.1

Dado $S \subset \mathbb{R}^n$, $d : S \times S \rightarrow \mathbb{R}$ llamado medida divergente en S si y sólo si

(i) $d(x, y) \geq 0$ para todo $x, y \in S$

(ii) Si $\{x_k\} \subset S, x \in S$ entonces $\lim_{k \rightarrow \infty} d(x, x^k) = 0$ si y solo si $\lim_{k \rightarrow \infty} x^k = x$

(iii) Los conjuntos de nivel $\Gamma_1(y, v) = \{x \in S : d(x, y) \leq v\}$ son acotados para todo $y \in S$ y todo $v > 0$.

(iv) Los conjuntos de nivel $\Gamma_2(x, v) = \{y \in S : d(x, y) \leq v\}$ son acotados.

Se introduce una familia de medidas divergentes definidas sobre el ortante positivo

$$S = \mathbb{R}_{++}^n : \{x \in \mathbb{R}^n : x > 0\}.$$

Sea G el conjunto de funciones continuas $\Gamma : \mathbb{R}_{++} \rightarrow \mathbb{R}$, las cuales satisfacen:

(a) φ es decreciente en $(0, 1)$ y creciente en $(1, +\infty)$

(b) $\varphi(1) = 0$

$$(c) \lim_{k \rightarrow \infty} \varphi(t) = +\infty$$

Definición 2.7.2

Sea $\varphi \in G$, se define $d_\varphi : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ como

$$d_\varphi(x, y) = \sum_{j=1}^n y_j \varphi\left(\frac{x_j}{y_j}\right)$$

la cual es llamada φ -divergencia.

Proposición 2.7.1 (Ver Prop. 2.1, [11])

Si $\varphi \in G$, entonces d_φ es una medida divergente en \mathbb{R}_{++}^n .

Para el estudio del análisis de convergencia, se considerarán unas condiciones adicionales al conjunto G .

Así se introduce un subconjunto Φ de G .

Sea Φ una clase de funciones estrictamente convexas y tres veces continuamente diferenciable $\varphi : \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ definidas por

$$\Phi : \{\varphi \in C^3(0, +\infty) : \varphi(1) = \varphi'(1) = 0, \varphi''(1) > 0\}$$

Proposición 2.7.2 (Ver Prop. 2.2, [11])

$$\Phi \subset G$$

Corolario 2.7.1 (ver Coro. 2.1, [11])

Si $\varphi \in \Phi$, entonces d_φ es una medida divergente.

Algunos ejemplos clásicos de φ -divergencias, son los siguientes:

Ejemplo 2.7.1

Si $\psi(t) = t \log t - t + 1$, entonces $d_\psi(x, y) = \sum_{j=1}^n (x_j \log \frac{x_j}{y_j} + y_j - x_j)$.

Obsérvese que esta es la casi-distancia Kullback-Liebler, la cual fue obtenida anteriormente vía distancia de Bregman.

Ejemplo 2.7.2

Si $\varphi(t) = (\sqrt{t} - 1)^2$, entonces $d_\varphi(x, y) = \sum_{j=1}^n (\sqrt{x_j} - \sqrt{y_j})^2$

Definición 2.7.3

Sea $\varphi \in \Phi$ defínase $\varphi^\diamond, \widehat{\varphi} : \mathbb{R}_{++} \rightarrow \mathbb{R}$ dadas por

$$\varphi^\diamond(t) = t\varphi\left(\frac{1}{t}\right) \quad (2.38)$$

$$\widehat{\varphi}(t) = (t - 1)\varphi'(t) \quad (2.39)$$

La función (2.38) fue introducida por Bental, Ben-Israel y Teboulle (1991) y es llamada la adjunta de φ .

Algunas propiedades de $\varphi^\diamond, \widehat{\varphi}$ y su relación con φ se establecen en el siguiente teorema.

Proposición 2.7.3 (ver Prop. 2.3, [11])

Sea $\varphi \in \Phi$, entonces

(i) $\varphi^\diamond \in \Phi$

(ii) $\varphi^{\diamond\diamond} = \varphi$

$$(iii) \quad d_\varphi(x, y) = d_{\varphi^\diamond}(y, x)$$

$$(iv) \quad \varphi(t) \leq \widehat{\varphi}(t) \text{ para todo } t > 0$$

$$(v) \quad \widehat{\varphi} \in G$$

Corolario 2.7.2 (Ver Coro. 2.2, [11])

Si $\varphi \in \Phi$ entonces d_{φ^\diamond} y $d_{\widehat{\varphi}}$ son medidas divergentes.

Proposición 2.7.4 (Ver Prop. 2.5, [11])

Sea $\varphi \in \Phi$ y denótese $\alpha = [\varphi''(1)]^{-1} > 0$ entonces existe $\mu > 0$ tal que para todo $t \in \mathbb{R}_{++}$

$$(t - 1) - \alpha t\varphi'(t) \leq \mu\widehat{\varphi}(t) \quad \alpha t\varphi'(t) - t \log t \leq \mu\widehat{\varphi}(t)$$

Con las φ -divergencias se ha desarrollado un método de punto proximal, el cual se presenta a continuación.

Sea U un conjunto convexo, $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ una función continua y convexa, y considerese el problema de minimización convexa

$$(P) \quad \min\{f(x) : x \in \mathbb{R}_+^n\} \tag{2.40}$$

Como se mencionó anteriormente, se reemplaza la distancia cuadrática usada en el algoritmo de minimización proximal clásico por φ -divergencia, y se introduce el Entropy-like Proximal Method (EPM) para resolver (P):

$$x_0 \in \mathbb{R}_{++}^n \tag{2.41}$$

$$x^{k+1} = \underset{x \in \mathbb{R}_+^n}{\operatorname{argmin}}\{f(x) + \lambda_k d_\varphi(x, x^k)\} \tag{2.42}$$

donde $\varphi \in \Phi_1 = \{\varphi \in \Phi : \lim_{t \rightarrow 0} \varphi'(t) = -\infty\}$ y $\lambda_k \in [\widehat{\lambda}, \bar{\lambda}] \subset (0, +\infty)$

Denótese por T el conjunto de soluciones de (P), asumiremos que $T \neq \emptyset$ y $\partial f(x)$ es no vacío para todo $x \in U$.

Definición 2.7.4

Defínase $u^k \in \mathbb{R}^n$ como

$$u_j^k = -\lambda_k \varphi' \left(\frac{x_j^{k+1}}{x_j^k} \right) \quad j = 1, \dots, n \quad (2.43)$$

Proposición 2.7.5 (ver Prop. 4.1, [11])

La sucesión generada para el algoritmo (2.40)-(2.41) está bien definida y está contenida en \mathbb{R}_{++}^n . Más aún $u^k \in \partial f(x^{k+1})$ para todo $k \geq 0$.

Proposición 2.7.6 (ver Prop. 4.2, [11])

Para todo $k \geq 0$

$$0 \leq \lambda_k d_\varphi(x^{k+1}, x^k) \leq \lambda_k d_{\widehat{\varphi}}(x^{k+1}, x^k) \leq f(x^k) - f(x^{k+1})$$

Corolario 2.7.3 (Ver Coro. 4.1, [11])

- (i) La sucesión $\{f(x^k)\}$ es decreciente y convergente
- (ii) $\sum_{k=0}^{\infty} \lambda_k d_{\widehat{\varphi}}(x^{k+1}, x^k) < \infty$
- (iii) $\lim_{t \rightarrow \infty} x_j \varphi(x_j^{k+1}/x_j^k) = 0$ para todo $j = 1, \dots, n$

Se introducirá la siguiente notación, para $x \in \mathbb{R}_+^n$ defínase $I(x) = \{j \in [1, n] : x_j = 0\}$, $J(x) = \{j \in [1, n] : x_j > 0\}$ y $j = 1, \dots, n$, $t_j^k = x_j^{k+1}/x_j^k$.

Proposición 2.7.7 (Ver Prop. 4.3, [11])

Si la sucesión $\{x^k\}$ converge a un punto x^* y

(i) f es continuamente diferenciable en \mathbb{R}_+^n , ó

(ii) Existe $z \in T$ tal que $I(x^*) \subset I(z)$

Entonces x^* pertenece a T .

Definición 2.7.5

Una sucesión $\{y^k\} \subset S$ es casi-fejér convergente a un conjunto V con respecto a la unidad de medida divergente d si para cada $v \in V$ existe $k \geq 0$ y una sucesión de número real $\epsilon_k \geq 0$ tal que $\sum_{k=0}^{\infty} \epsilon_k < \infty$ y

$$d(v, y^{k+1}) \leq d(v, y^k) + \epsilon_k \quad \text{para todo } k \geq K$$

Teorema 2.7.1 (Ver Teo. 4.1, [11]).

(i) Si una sucesión $\{y^k\}$ es quasi-Fejer convergente a un conjunto no vacío V , entonces $\{y^k\}$ es acotado y $\{d(v, y^k)\}$ es acotado para $v \in V$.

(ii) Si V contiene el punto límite de $\{y^k\}$, entonces $\{y^k\}$ converge.

Defínase $T^* = \{x \in U : f(x) \leq r\}$ claramente $T^* \supset T$ donde $r = \lim_{k \rightarrow \infty} f(x^k)$ el cual existe para el colorario (2.7.3) y sea T_1 y T_2 , donde T_1 es el conjunto de punto límite de $\{x^k\}$ por el corolario (2.7.3) (i) $T_1 \subset T^*$. Sea I el conjunto de índices y del cual existe una subsucesión $\{x_j^{l_k}\}$ de $\{x_j^k\}$ la cual converge a 0. T_2 está definida como

$$T_2 = \{x \in T^* : I \subset (x)\}, \text{ i.e } T_2$$

En la siguiente proposición se tomará ψ como en el ejemplo (2.7.1).

Proposición 2.7.8 (Ver Prop. 4.4, [11])

Sea cualquier $z \in T^*$ y μ como en **Proposición (2.7.4)**, entonces:

$$(i) \ d_\psi(z, x^{k+1}) - d_\psi(z, x^k) \leq \frac{\mu}{\lambda} \lambda_k \sum_{j=1}^n \left(1 + \frac{z_j}{x_j^{k+1}}\right) x_j^k \widehat{\varphi}(t_j^k) \text{ para todo } k \geq 0$$

(ii) Si $\varphi'(t) \leq \varphi''(1) \log t$ para todo $t > 0$. Entonces

$$d_\psi(z, x^{k+1}) - d_\psi(z, x^k) \leq \frac{\mu}{\lambda} \lambda_k d\widehat{\varphi}(x^k, x^{k+1}) \text{ para todo } k \geq 0$$

Corolario 2.7.4 (Ver Coro. 4.2, [11])

(i) $\{x^k\}$ es casi-Fejér convergente T_2 con respecto a d_ψ

(ii) Si $\varphi'(t) \leq \varphi''(1) \log t$ para todo $t > 0$ entonces $\{x^k\}$ es casi-fejér convergente a T^* con respecto a d_ψ .

Proposición 2.7.9 (Ver Prop. 4.5, [11])

Si existe k_0 tal que

$$f(x^{k_0}) - r \leq \min \left\{ \frac{r}{2\tilde{\mu}}, \widehat{\lambda} \xi \varphi \left(\frac{1}{2} \right) \right\} \quad (2.44)$$

$$d_\psi(z, x^{k_0}) \leq \frac{r}{2} \quad (2.45)$$

entonces, para todo $k \geq k_0$

$$d_\psi(z, x^{k+1}) - d_\psi(z, x^k) \leq \tilde{\mu} \lambda_k d\widehat{\varphi}(x^k, x^{k+1})$$

Corolario 2.7.5 (Ver Coro. 4.3, [11])

- (i) $\{x^k\}$ es cuasi-fejér convergente en T , con respecto a d_ψ .
- (ii) Si x^0 es suficientemente cercano a un punto $z^* \in T$ entonces $\{x^k\}$ es casi-fejér convergente a $\{z^*\}$ con respecto a d_ψ .

Finalmente se sumarán todos los resultados en el siguiente teorema, armado con los resultados del Corolario (2.7.4) y Corolario (2.7.5) y el Teorema (2.7.1) y la Proposición (2.7.7).

Teorema 2.7.2 (Ver Teo. 4.2 [11]).

Si cualquiera de las siguientes condiciones se tiene:

- (i) La sucesión $\{x^k\}$ tiene puntos límites.
- (ii) El conjunto T es acotado.
- (iii) Existe una solución z del problema (P) tal que $\{x_j^k\}$ es acotada para aquellos componentes tales que $z_j > 0$
- (iv) $\varphi'(t) \leq \varphi''(1) \log t$ para todo $t > 0$.
- (v) Si x^0 es suficientemente cercano a la solución z^* del problema (2.39), entonces la sucesión $\{x^k\}$ generada por (2.41) converge.

Más aún, bajo cualquiera de las condiciones (iii), (iv) o (v) o si f es $C'(\mathbb{R}_+^n)$ cualquier punto límite de la sucesión es una solución del problema (2.39).

Corolario 2.7.6 (Ver Coro. 4.4, [11])

Si $\varphi'(t) \leq \varphi''(1) \log t$, para todo $t > 0$, entonces la sucesión generada por (2.41) converge a una solución del problema (2.39).

Obsérvese que la distancia entrópica Kullback-Leibler puede ser obtenida tanto de las φ -divergencias como las distancias de Bregman. Es de hacer notar que en este trabajo se ofrece otra conexión entre ambos métodos, la cual será destacada más adelante.

Capítulo 3

Una Casi-distancia Generalizada en los casos $0 < p < 2$

En este capítulo se presentan los aportes originales de este trabajo, en la sección 3.1 se hace una breve introducción a la casi-distancia generalizada, propuesta en [19]. En la sección 3.2 se propone una familia de penalidad, que a diferencia de [11] no pasa por el origen con pendiente uno. En la sección 3.3, se presenta el algoritmo de punto proximal con la casi-distancia generalizada, caso $0 < p < 2$, esta sección se divide en tres partes, la primera subsección se presentan los teoremas y propiedades básicos para obtener la convergencia siguiendo los pasos de [11] y [19], en la siguiente subsección se presenta la noción de casi-Fejer convergencia y por último en la tercera subsección se realizan los estudios de convergencia desde el punto de vista dual-primal. Para finalizar en la última sección se presenta la convergencia desde el punto de vista primal-dual.

3.1. Introducción a la casi-distancia generalizada.

El problema de programación convexa es expresado comunmente de la siguiente manera:

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

donde las funciones $f_i : U \subset \mathbb{R}_+^n \rightarrow \mathbb{R}$, para $i = 0, 1, \dots, n$ son convexas, propias y cerradas. El problema (P) ha sido estudiado y resuelto, tanto por medio de métodos teóricos como métodos computacionales, debido a su aplicabilidad en múltiples problemas de ingeniería y en sus diversas áreas, los cuales pueden ser modelados bajo este formato.

Asociado al problema (P) se tiene el problema dual, el cual viene dado por:

$$d^* = \sup\{d(\mu) : \mu \geq 0\} \quad (D) \quad (3.2)$$

donde,

$$d(\mu) = \inf\{l(x, \mu) : x \in \mathbb{R}^n\} \quad (3.3)$$

y

$$l(x, \mu) = f_0(x) + \sum_{i=1}^n \mu_i f_i(x) \quad (3.4)$$

es la función Lagrangeana.

Suponga las siguientes hipótesis:

(H1) El conjunto de soluciones de (P) es no vacío y compacto.

(H2) Existe \bar{x} tal que $f_i(\bar{x}) < 0$, para $i = 1, \dots, m$. (C. Slater).

Una forma de resolver el problema (P) consiste en utilizar los métodos de Lagrangeano aumentado, los cuales, debido a su eficiente estabilidad computacional en una diversidad de problemas y su conocida relación teórica con los métodos de punto proximal, via dualidad de Fenchel, han sido estudiados ampliamente en los últimos años. Es bien conocido que los métodos de Lagrangeano aumentado pueden ser introducidos desde un punto de vista primal, ver [2], [3] y [15] o desde el punto de vista dual, donde el método de multiplicadores es construido aplicando teoría de dualidad de Fenchel al método de

punto proximal, ver [1], [6] y [11].

En la sección 3.2, se estudiará una casi-distancia generalizada introducida por [19] para resolver el problema (D) la cual esta dada por la siguiente expresión:

$$d_{\varphi}^p(x, y) = \sum_{i=1}^n \left[\frac{y_i^p}{c} \varphi\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c} \varphi(c) - y_i^{p-1} \varphi'(c)(x_i - y_i) \right] \quad (3.5)$$

donde φ es una función que no pasa a través del origen y satisface que $\varphi'(0) = k, k > 0$ con $p \geq 0$ y $c > 0$ es tal que para un $\tilde{y} \in \mathbb{R}$ se tiene que $\varphi'(\tilde{y}) = c$ y $d_{\varphi}^p : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}$, en la sección 3.3, se presenta el problema original continuando con los problemas abiertos dejados en [19], por último, en la sección 3.4 se desarrollarán técnicas similares a las desarrolladas en [11] para los casos $0 < p < 1$ y $p = 1$ generalizados, basados en un algoritmo de tipo punto proximal y un algoritmo de Lagrangeano aumentado, en el estudio de φ -divergencias trasladadas mostrando resultados en los espacios primal y dual.

3.2. Familia de Penalidad propuesta.

La familia de penalidad que se introducirá esta definida sobre el ortante positivo

$$S := \mathbb{R}_{++}^n = \{x \in \mathbb{R}^n : x > 0\}$$

Sea Φ el conjunto de funciones de clase C^3 , estrictamente convexas $\varphi : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ donde φ es la función conjugada asociada a la función de penalidad la cual satisface:

1. $\varphi(\bar{k}) = \min\{\varphi(t) : 0 < t < +\infty\}$
2. φ es decreciente en $(0, \bar{k})$ y creciente en $(\bar{k}, +\infty)$
3. $\lim_{t \rightarrow \infty} \varphi'(t) = +\infty$
4. $\lim_{t \rightarrow 0} \varphi'(t) = -\infty$
5. $\varphi''(t) \leq M$ para todo $t \geq \bar{k}$

$$6. \varphi''(t) > (\bar{k} - t)\varphi'''(t), \forall t > 0$$

A diferencia de [11] la función usada no necesita pasar por el origen con pendiente uno y a diferencia de [19], se introduce la propiedad 3, y la segunda derivada de la función de penalidad no es acotada superiormente.

Ejemplo 3.2.1 Para $\theta(t) = ke^t - 1$ (véase la figura 3.1), se tiene que $\theta(0) = k - 1$, $\theta'(0) = k$ y $\theta^*(t) = \varphi(t) = t \ln(t) - t - t \ln(k) + 1$ (véase la figura 3.2), donde $\varphi'(k) = 0$ y $\varphi(k) = -k + 1$

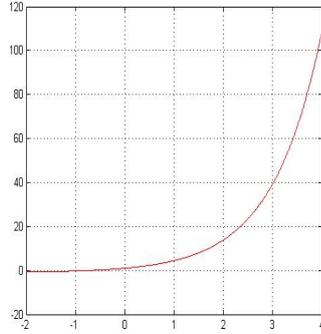


Figura 3.1: Función θ

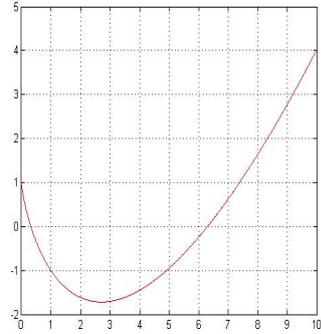


Figura 3.2: Función θ^*

De acuerdo a la **Definición 2.7.1**, definase la casi-distancia como:

$$d_\varphi^p(x, y) = \sum_{i=1}^n \left[\frac{y_i^p}{c} \varphi\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c} \varphi(c) - y_i^{p-1} \varphi'(c)(x_i - y_i) \right] \quad (3.6)$$

En la siguiente proposición se probará que (3.6) es en efecto una medida divergente, pero observese antes algunos detalles importantes. Note que $d_\varphi^p(\cdot, y)$ es una función estrictamente convexa porque φ lo es. Al particularizar para los valores $p = 1$, $c = 1$, $\tilde{y} = 0$ con $\theta(\tilde{y}) = 0$ se obtiene:

$$P_1^*(s, \mu, r, c) = rd_{\theta^*}(s, \mu) = r \sum_{i=1}^m \mu_i \theta^*\left(\frac{s_i}{\mu_i}\right) \quad (3.7)$$

la cual es usada en métodos de punto proximal con ϕ -divergencias, ver [11].

Para $p=2$, $c=1$, $\tilde{y} = 0$ con $\theta(\tilde{y}) = 0$ así,

$$P_2^*(s, \mu, r, c) = r\tilde{d}_{\theta^*}(s, \mu) = r \sum_{i=1}^m \mu_i^2 \theta^*\left(\frac{s_i}{\mu_i}\right) \quad (3.8)$$

obteniéndose los núcleos homogéneos de segundo orden usados por [3].

Ejemplo 3.2.2 (Una generalización de la distancia entrópica Kullback-Leibler)

Para $\varphi(t) = t \ln(t) - t$ se tiene que $d_\varphi^p(x, y) = y^{p-1}(x \ln(\frac{x}{y}) - x + y)$. Esta casi-distancia puede ser considerada una generalización de la distancia entrópica Kullback-Leibler la cual es muy usada en la Teoría de probabilidad y de la información y fue introducida originalmente por Solomon Kullback y Richard Leibler en 1951 .

Se probará ahora que la casi-distancia considerada es una medida divergente.

Proposición 3.2.1 (Ver Prop. 3.2.3.1, [19])

Si $\varphi \in \Phi$ entonces d_φ^p es medida divergente con $p \geq 0$ y $c > 0$.

En conexión con cualquier $\varphi \in \phi$ se define una función $\hat{\varphi} : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ dada por

$$\hat{\varphi}(t) = \varphi(\bar{k}) + (t - \bar{k})\varphi'(\bar{k}) \quad (3.9)$$

Proposición 3.2.2

Sea $\varphi \in \Phi$ entonces

1. $\hat{\varphi} \in \Phi$
2. $\varphi(t) \leq \hat{\varphi}(t)$ para todo $t > 0$

Demostración:

1. $\hat{\varphi}$ satisface las condiciones de Φ . (ver pág. 58)

En efecto; de la condición 5 se tiene que,

$$\varphi''(t) > (\bar{k} - t)\varphi'''(t), \quad \text{con } t \neq \bar{k}$$

$$2\varphi''(t) > (\bar{k} - t)\varphi'''(t),$$

$$2\varphi''(t) + (t - \bar{k})\varphi'''(t) > 0,$$

esto es,

$$\widehat{\varphi}''(t) > 0,$$

así $\widehat{\varphi}$ es estrictamente convexa.

Por otra parte

$\widehat{\varphi}'(\bar{k}) = \varphi'(\bar{k}) = 0$ y $\widehat{\varphi}(\bar{k}) = \varphi(\bar{k})$ así se satisface 2, además $\varphi''(t) > 0$ para todo t por lo que $\widehat{\varphi}' > 0$ para todo $t > \bar{k}$ y así $\widehat{\varphi}$ es creciente en (\bar{k}, ∞^+) y análogamente $\widehat{\varphi}$ es decreciente $(0, \bar{k})$, 3 y 4 se cumple inmediatamente, y 5 se satisface ya que $\widehat{\varphi}(t)$ es estrictamente convexa.

2. Se probará que $\varphi(t) \leq \widehat{\varphi}(t)$ para todo $t > 0$. De la convexidad de φ se tiene que:

$$\varphi(\bar{k}) \geq \varphi(t) + \varphi'(\bar{k})(\bar{k} - t), \quad \forall t \in \mathbb{R}_{++}$$

$$\varphi(\bar{k}) + \varphi'(\bar{k})(t - \bar{k}) \geq \varphi(t)$$

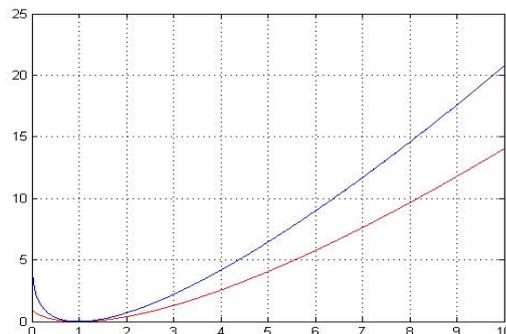
por tanto $\varphi(t) \leq \widehat{\varphi}(t)$

■

Ejemplo 3.2.3 Sea $\varphi(t) = \psi(t) = t \ln(t) - t - t \ln(\bar{k}) + 1$, donde

$\widehat{\varphi}(t) = \widehat{\psi}(t) = 1 - \bar{k} + (t - \bar{k})(\ln(t) - 1)$. Obsérvese esta función generaliza a la función tomada en [11] cuando $\bar{k} = 1$. Donde

$$d_{\varphi}^p(x, y) = \sum_{i=1}^n \left[\frac{y_i^p}{c} \varphi\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c} \varphi(c) - y_i^{p-1} \varphi'(c)(x_i - y_i) \right]$$

Figura 3.3: Función $\varphi(t) \leq \widehat{\varphi}(t)$

3.3. Algoritmo proximal con una Casi-distancia Generalizada.

Caso $0 < p < 2$

Sea U un conjunto convexo $f : U \subset \mathbb{R}_+^n \rightarrow \mathbb{R}$ una función continua, cerrada y propia y considerese el problema de minimización

$$(D) \quad \min\{f(x) : x \in \mathbb{R}^n\} \quad (3.10)$$

reemplazando la distancia cuadrática del algoritmo de punto proximal por la medida divergente

$$d_\varphi^p(x, y) = \sum_{i=1}^n \left[\frac{y_i^p}{c} \varphi\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c} \varphi(c) - y_i^{p-1} \varphi'(c)(x_i - y_i) \right] \quad (3.11)$$

con $0 < p < 2$ y $c > 0$ para resolver

$$x^0 \in \mathbb{R}_{++}^n \quad (3.12)$$

$$x^{k+1} = \arg \min_{x \in \mathbb{R}_+^n} \{f(x) + \lambda_k d_\varphi^p(x, x^k)\} \quad (3.13)$$

donde $\varphi \in \Phi$ y $\lambda_k \in [\widehat{\lambda}, \bar{\lambda}] \subset (0, \infty)$.

Se denotará por T el conjunto de soluciones de (D) suponiendo que $T \neq \emptyset$ y $\partial f(x)$ es el subdiferencial de f , no-vacio para todo $x \in U$

Al obtener la casi-distancia de forma natural, se uso un shift constante para definir una función de penalidad generalizada veáse [19] sección 3.2.2.

Las tecnicas usadas para desarrollar este estudio no resultan convenientes para todo $c > 0$, así convenientemente se tomo $c = \bar{k}$ luego por el **Corolario 23.5.1** en [16], se tiene que:

$$(\theta') = (\theta^*)' \quad (3.14)$$

y así

$$\theta'(0) = c = \bar{k} \Leftrightarrow 0 = (\theta^*)'(\bar{k}) \quad (3.15)$$

Ejemplo 3.3.1 Dada la función de penalidad $\theta(y) = \bar{k}e^y - 1$ (veáse la figura 3.1). De acuerdo a [19] la función de penalidad viene dada por

$$P_{\mu,i}(y_i, \mu_{i,r,c}) = r \frac{\mu_i^p}{c} \left[\theta\left(\frac{y_i}{\mu_i^{p-1}r} + \tilde{y}\right) - \theta(\tilde{y}) \right] \quad \text{para } i = 1, \dots, m. \quad (3.16)$$

así

$$P_{\mu,i}(y_i, \mu_{i,r,c}) = r \frac{\mu_i^p}{c} \left[\bar{k}e^{\frac{y_i}{\mu_i^{p-1}r} + \tilde{y}} - \bar{k}e^{\tilde{y}} \right] \quad (3.17)$$

Observando que

$$(P_{\mu,i})'(0, \mu_{i,r,c}) = \mu_i, \quad \text{para } i = 1, \dots, m \quad (3.18)$$

por lo que al derivar (3.17) con respecto a y nos queda

$$(P_{\mu,i})'(y_i, \mu_{i,r,c}) = \frac{\mu_i}{c} \left[\bar{k}e^{\left(\frac{y_i}{\mu_i^{p-1}r} + \tilde{y}\right)} \right] \quad (3.19)$$

$$(P_{\mu,i})'(0, \mu_{i,r,c}) = \frac{\mu_i}{c} \left[\bar{k}e^{\tilde{y}} \right] \quad (3.20)$$

igualando (3.20) a μ_i se tiene que

$$\left[\bar{k}e^{\tilde{y}} \right] = c \quad (3.21)$$

para $\bar{k} = c$ se tiene que

$$\tilde{y} = \ln(1) = 0 \quad (3.22)$$

La casi-distancia generalizada obtenida para $c = \bar{k}$ es

$$d_{\varphi}^p(x, y) = \sum_{i=1}^n \left[\frac{y_i^p}{c} \varphi\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c} \varphi(c) \right] \quad (3.23)$$

Observese que (3.23) generaliza el caso $p = c = \bar{k} = 1$ el cual fue estudiado en [11].

Vale la pena destacar que las casi-distancias asociadas a φ -divergencias y distancias de Bregman no guardan una estrecha relación entre si y las pruebas de convergencia en cada caso son muy diferentes, sin embargo, como puede observarse en [11], se puede obtener la misma casi-distancia a partir de una función de Bregman adecuada y una φ -divergencia particular, aunque, en [19] ambas φ -divergencia y Distancias de Bregman están relacionadas mediante la casi-distancia generalizada como un caso particulares de esta. En [19] se obtuvo la convergencia para el caso $p \geq 2$.

El aporte principal del presente trabajo es abrir caminos para las casi-distancias generalizadas obteniendo la convergencia para algunos valores de p que permanecían abiertos, caso ($0 < p < 2$). Sin embargo, las tecnicas utilizadas en [11] no resultan convenientes para obtener resultados de convergencia para todo $c > 0$, por problemas de no acotabilidad, pero en el caso particular de $c = \bar{k}$ si se pudo obtener dichos resultados, lo que sugiere que el shift constante a considerar debe ser igual al minimo de la función de penalidad trasladada.

En la siguiente sección se presenta el analisis de convergencia para dichos casos.

3.4. Análisis de convergencia.

3.4.1. Resultados Preliminares.

Observación 3.4.1 Definamos $u_j^k \in \mathbb{R}^n$ como

$$u_j^k = -\lambda_k((x_j^k)^{p-1} \varphi'(\frac{cx_j^{k+1}}{x_j^k})) \quad j = 1, \dots, n \quad (3.24)$$

Proposición 3.4.1 (Ver Prop. 4.1, [11])

La sucesión generada por (3.12) y (3.13) está bien definida y contenida en \mathbb{R}_{++}^n .
Más aun $u^k \in \partial f(x^{k+1})$ para todo $k \geq 0$

Demostración:

Procediendo por inducción

Para $k = 0$ se cumple que $x^0 \in \mathbb{R}_{++}^n$ por (3.12)

La prueba consiste en demostrar que x^{k+1} existe.

f es continua en U y T es el conjunto de minimizadores de (P) con $T \neq \emptyset$, así, f es acotada inferiormente y $\varphi \in \Phi$ donde d_φ^p es una medida divergente, luego por la definición de medida divergente, los conjuntos de nivel de $d_\varphi^p(x, y)$ son acotados y así el minimizador está en (3.13) reduciendo el estudio a un conjunto compacto y f_k logra el mínimo, donde $f_k(x) = f(x) + \lambda_k d_\varphi^p(x, y)$. Puesto que φ es estrictamente convexa, el minimizador es único, lo que significa que x^{k+1} esta bien definida.

Se demostrara ahora que $x^{k+1} > 0$.

Puesto que f_k es convexa la condición de optimalidad de (3.13) implica que,

$$0 \in \partial(f_k(x) + \delta(x/\mathbb{R}_+^n))$$

$$0 \in \partial f_k(x) + \partial \delta(x/\mathbb{R}_+^n)$$

$$0 \in \partial(f(x) + \lambda_k d_\varphi^p(x^{k+1}, x^k)) + \partial \delta(x/\mathbb{R}_+^n)$$

$$\begin{aligned}
0 \in \partial f(x^{k+1}) + \lambda_k (x_i^k)^p \varphi' \left(\frac{cx_i^{k+1}}{x_i^k} \right) + \partial \delta(x/\mathbb{R}_+^n) \quad i = 1, \dots, n \\
0 \in \partial f(x^{k+1}) - u^k - v^k, \quad -v^k \in \partial \delta(x^{k+1}/\mathbb{R}_+^n)
\end{aligned} \tag{3.25}$$

donde $\delta(\bullet/\mathbb{R}_+^n)$ denota la función indicadora, dada por $\delta(x/\mathbb{R}_+^n) = \begin{cases} 0 & x \in \mathbb{R}_+^n \\ +\infty & x \notin \mathbb{R}_+^n \end{cases}$;

la primera inclusión de (3.24) implica que u^k es finito, puesto que $\varphi \in \Phi$ y $\lim_{t \rightarrow 0} \varphi'(t) = -\infty$ y así x^{k+1} , no puede ser igual a cero, por lo que se concluye que $x^{k+1} > 0$. Finalmente, de la segunda inclusión de (3.25) se tiene que $v^k = 0$ y así por la primera inclusión de (3.25), $u_j^k \in \partial f(x^{k+1})$ para todo i . ■

Proposición 3.4.2 (Ver Prop. 4.2, [11])

Para $\bar{k} = c$ y $0 < p < 2$, se tiene

$$0 \leq \lambda_k d_\varphi^p(x^{k+1}, x^k) \leq \lambda_k d_{\hat{\varphi}}^p(x^{k+1}, x^k) \leq f(x^k) - f(x^{k+1})$$

Demostración:

Observando (3.6) y ya que $c = \bar{k}$,

$$d_\varphi^p(x^{k+1}, x^k) = \sum_{i=1}^n \left[\frac{(x_i^k)^p}{c} \varphi \left(\frac{cx_i^{k+1}}{x_i^k} \right) - \frac{(x_i^k)^p}{c} \varphi(c) \right], \tag{3.26}$$

por otro lado $d_{\hat{\varphi}}^p$ toma la forma,

$$d_{\hat{\varphi}}^p(x^{k+1}, x^k) = \sum_{i=1}^n \left[(x_i^k)^{p-1} (x_i^{k+1} - x_i^k) \varphi' \left(\frac{cx_i^{k+1}}{x_i^k} \right) \right], \tag{3.27}$$

de la **Proposición 3.2.2**, se tiene que, $\varphi(t) \leq \hat{\varphi}(t)$ para todo $t > 0$ en particular para $t = \frac{cx_i^{k+1}}{x_i^k}$ con $x_i^{k+1} > 0$ y $x_i^k > 0$ se tiene que,

$$\varphi \left(\frac{cx_i^{k+1}}{x_i^k} \right) \leq \varphi(\bar{k}) + \left(\frac{cx_i^{k+1}}{x_i^k} - \bar{k} \right) \varphi' \left(\frac{cx_i^{k+1}}{x_i^k} \right),$$

ya que $\frac{(x_i^k)^p}{c} > 0$ multiplicando en ambos lados de la desigualdad para $i = 1, \dots, n$

$$\frac{(x_i^k)^p}{c} \varphi \left(\frac{cx_i^{k+1}}{x_i^k} \right) \leq \frac{(x_i^k)^p}{c} \varphi(\bar{k}) + \frac{(x_i^k)^p}{c} \left(\frac{cx_i^{k+1}}{x_i^k} - \bar{k} \right) \varphi' \left(\frac{cx_i^{k+1}}{x_i^k} \right),$$

así, aplicando la desigualdad anterior en (3.27) se tiene que

$$d_{\varphi}^p(x^{k+1}, x^k) \leq \sum_{i=1}^n [(x_i^k)^{p-1} (x_i^{k+1} - x_i^k) \varphi'(\frac{cx_i^{k+1}}{x_i^k})] = d_{\varphi}^p(x^{k+1}, x^k),$$

multiplicando en ambos lados de la desigualdad por λ_k se obtiene la desigualdad deseada.

Por último, ya que $\lambda_k d_{\varphi}^p(x^{k+1}, x^k) = \lambda_k \sum_{i=1}^n [(x_i^k)^{p-1} (x_i^{k+1} - x_i^k) \varphi'(\frac{cx_i^{k+1}}{x_i^k})] = (x^k - x^{k+1})u^k$ por la **Proposición 3.2.1** $u^k \in \partial f(x^{k+1})$ y así la conclusión se sigue de la definición de subgradiente. ■

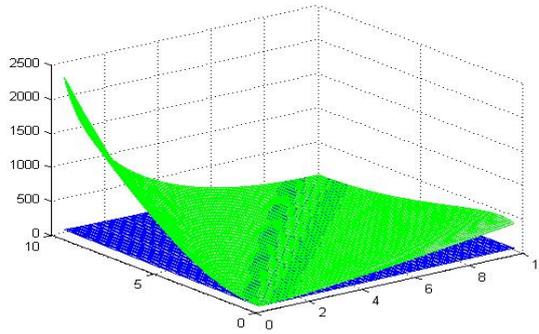


Figura 3.4: $c = \bar{k} = 20, p = 1,6, d_{\varphi} \geq d\varphi$

Corolario 3.4.1 (Ver Coro. 4.1, [11])

1. La sucesión $\{f(x^k)\}$ es decreciente y convergente.
2. $\sum_{i=1}^{\infty} \lambda_k d_{\varphi}^p(x^{k+1}, x^k) < \infty$

$$3. \lim_{k \rightarrow \infty} \frac{(x_i^k)^p}{c} \varphi\left(\frac{cx_i^{k+1}}{x_i^k}\right) - \frac{(x_i^k)^p}{c} \varphi(c) = 0 \quad \text{para todo } i = 1, \dots, n$$

Demostración:

1. La sucesión $\{f(x^k)\}$ es decreciente debido la **Proposición 3.4.2** y acotada inferiormente ya que el conjunto de soluciones T es no-vacío así, la sucesión es convergente.
2. De la **Proposición 3.2.2** se tiene que

$$0 \leq \sum_{i=1}^k d_{\varphi}^p(x^{i+1}, x^i) \leq f(x^0) - f(x^{k+1}) \leq f(x^0) - f(z) \quad \text{para todo } i = 1, \dots, n \quad \text{y } z \in T,$$

por lo que

$$\sum_{k=0}^{\infty} \lambda_k d_{\varphi}^p(x^{k+1}, x^k) \leq f(x^0) - f(x^{k+1}) < \infty$$

3. Esto es inmediato ya que por (2), se tiene que $\lim_{k \rightarrow \infty} \lambda_k d_{\varphi}(x^{i+1}, x^i) = 0$, puesto que $\lambda_k > 0$

$$\lim_{k \rightarrow \infty} \sum_{i=1}^n \left[\frac{(x_i^k)^p}{c} \varphi\left(\frac{cx_i^{k+1}}{x_i^k}\right) - \frac{(x_i^k)^p}{c} \varphi(c) \right] = 0$$

para todo $i = 1, \dots, n$ se tiene que $\lim_{k \rightarrow \infty} \frac{(x_i^k)^p}{c} \varphi\left(\frac{cx_i^{k+1}}{x_i^k}\right) - \frac{(x_i^k)^p}{c} \varphi(c) = 0$

■

Proposición 3.4.3

La sucesión generada por (3.12) y (3.13) es acotada.

Demostración:

Por (H2) y ya que f es una función propia convexa el conjunto de nivel $\Lambda = \{x \in \mathbb{R}^n / f(x^k) \leq f(x^0)\}$ es compacto y por el **Colorario 3.4.1**, $x_j^k \in \Lambda$ para todo k , así x_j^k es acotada.

■

Considerese la siguiente notación.

Para $x \in \mathbb{R}_+^n$, definase $I(x) = \{j \in [1, n] : x_j = 0\}$, $J(x) = \{j \in [1, n] : x_j > 0\}$ y $t_j^k = \frac{x_j^{k+1}}{x_j^k}$ para todo $j = 1, \dots, n$.

Proposición 3.4.4 (Ver Prop. 4.3, [11])

Si la sucesión $\{x^k\}$ converge al punto x^* y

1. f es continuamente diferenciable en \mathbb{R}_+^n , o
2. Existe $z \in T$ tal que $I(x^*) \subset I(z)$

entonces x^* pertenece a T

Demostración:

Tomando $z \in T$. Por la **Proposición 3.2.2** y de la definición de subgradiente

$$f(x^{k+1}) - f(z) \leq (x^{k+1} - z)^T u^k$$

$$= \sum_{j \in J(x^*)} (x_j^{k+1} - z_j) u_j^k + \sum_{j \in I(x^k)} x_j^{k+1} u_j^k - \sum_{j \in I(x^*)} z_j u_j^k \quad (3.28)$$

Se probará que el primer y segundo sumando de (3.28) tiende a cero cuando $k \rightarrow \infty$

Para $j \in J(x^*)$. Por el **Colorario 3.4.1** se tiene

$$0 = \lim_{k \rightarrow \infty} \frac{(x_j^k)^p}{c} \varphi\left(\frac{cx_j^{k+1}}{x_j^k}\right) - \frac{(x_j^k)^p}{c} \varphi(c) \quad \text{para todo } j = 1, \dots, n \quad (3.29)$$

$$0 = \lim_{k \rightarrow \infty} \frac{(x_j^k)^p}{c} [\varphi\left(\frac{cx_j^{k+1}}{x_j^k}\right) - \varphi(c)] = \frac{(x_j^*)^p}{c} \lim_{k \rightarrow \infty} [\varphi\left(\frac{cx_j^{k+1}}{x_j^k}\right) - \varphi(c)], \quad (3.30)$$

de (3.30) se tiene que $\lim_{k \rightarrow \infty} ct_j^k = c$, como $\varphi \in \Phi$ y $\lim_{k \rightarrow \infty} t_j^k = 1$, entonces $c = \lim_{k \rightarrow \infty} ct_j^k$, y así $0 = \varphi'(c) = \varphi'(\lim_{k \rightarrow \infty} ct_j^k) = \lim_{k \rightarrow \infty} \varphi'(ct_j^k)$

$$0 = - \lim_{k \rightarrow \infty} \varphi'(ct_j^k),$$

multiplicado en ambos lados de la desigualdad por $(x_j^*)^{p-1}$ se tiene que

$$0 = - \lim_{k \rightarrow \infty} (x_j^k)^{p-1} \varphi'(ct_j^k) = \lim_{k \rightarrow \infty} \frac{u_j^k}{\lambda_k},$$

puesto que $\lambda_k \leq \bar{\lambda}$, para todo k , se sigue que $\lim_{k \rightarrow \infty} u_j^k = 0$, implicando

$$\lim_{k \rightarrow \infty} \sum_{j \in J(x^*)} (x_j^{k+1} - z_j) u_j^k = 0. \quad (3.31)$$

Estudiese el segundo término de (3.28) de la siguiente manera

Si $\frac{1}{2} \leq t_j^k \leq 2$ se tiene que;

$$|x_j^{k+1}u_j^k| \leq \bar{\lambda} \max\{-\varphi'(\frac{c}{2})(x_j^k)^{p-1}, \varphi'(2c)(x_j^k)^{p-1}\}x_j^{k+1}, \quad (3.32)$$

puesto φ' es estrictamente creciente.

Estudiese ahora los casos $t_j^k < \frac{1}{2}$ y $t_j^k > 2$,

para

$$t_j^k < \frac{1}{2} \Rightarrow \frac{x_j^{k+1}}{x_j^k} < \frac{1}{2} \Rightarrow 2x_j^{k+1} < x_j^k \Rightarrow x_j^{k+1} < x_j^k - x_j^{k+1},$$

por otra parte

$$t_j^k < \frac{1}{2} < 1 \Rightarrow t_j^k < 1 \Rightarrow ct_j^k < c \Rightarrow \varphi'(ct_j^k) < \varphi'(c) = 0,$$

así $\varphi'(ct_j^k) < 0$, por lo que

$$u_j^k = -\lambda_k(x_j^k)^{p-1}\varphi'\left(\frac{cx_j^{k+1}}{x_j^k}\right) > 0,$$

así

$$\begin{aligned} 0 \leq x_j^{k+1}u_j^k &\Rightarrow 0 \leq -\lambda_k x_j^{k+1}(x_j^k)^{p-1}\varphi'(ct_j^k) < -\lambda_k((x_j^k) - (x_j^{k+1}))(x_j^k)^{p-1}\varphi'(ct_j^k) \\ &= \lambda_k((x_j^{k+1}) - (x_j^k))(x_j^k)^{p-1}\varphi'(ct_j^k), \end{aligned}$$

de aquí

$$|x_j^{k+1}u_j^k| < \lambda_k d_{\varphi}^p(x^{k+1}, x^k). \quad (3.33)$$

Estudiando ahora el caso en que $t_j^k > 2$

$$\frac{x_j^{k+1}}{x_j^k} > 2 \Rightarrow x_j^{k+1} > 2x_j^k \Rightarrow -x_j^{k+1} < -2x_j^k \Rightarrow 2x_j^{k+1} - x_j^{k+1} < x_j^{k+1} - 2x_j^k \Rightarrow x_j^{k+1} < 2(x_j^{k+1} - x_j^k),$$

por otro lado

$$t_j^k > 2 > 1 \Rightarrow ct_j^k > c \Rightarrow \varphi'(ct_j^k) > \varphi'(c) = 0,$$

por lo que $\varphi'(ct_j^k) > 0$,

así

$$u_j^k = -\lambda_k(x_j^k)^{p-1}\varphi'\left(\frac{cx_j^{k+1}}{x_j^k}\right) < 0,$$

luego

$$0 \leq -x_j^{k+1}u_j^k \Rightarrow 0 \leq \lambda_k x_j^{k+1}(x_j^k)^{p-1}\varphi'(ct_j^k) < 2\lambda_k(x_j^{k+1} - x_j^k)(x_j^k)^{p-1}\varphi'(ct_j^k),$$

de aquí

$$|x_j^{k+1}u_j^k| < 2\lambda_k d_{\varphi^p}^p(x^{k+1}, x^k). \quad (3.34)$$

De (3.31), (3.33) y (3.34) se obtiene

$$|x_j^{k+1}u_j^k| \leq \max\{2\lambda_k d_{\varphi^p}^p(x^{k+1}, x^k), \bar{\lambda} \max\{-\varphi'\left(\frac{c}{2}\right)(x_j^k)^{p-1}, \varphi'(2c)(x_j^k)^{p-1}\}x_j^{k+1}\}. \quad (3.35)$$

Puesto que $\lim_{k \rightarrow \infty} \lambda_k d_{\varphi^p}^p(x^{k+1}, x^k) = 0$, por el **Colorario 3.4.1** inciso 2, y $\lim_{k \rightarrow \infty} x_j^{k+1} = x_j^* = 0$, porque $j \in I(x^*)$, y así se concluye de (3.35) que $\lim_{k \rightarrow \infty} x_j^{k+1}u_j^k = 0$, así

$$\lim_{k \rightarrow \infty} \sum_{j \in I(x^*)} x_j^{k+1}u_j^k = 0. \quad (3.36)$$

Ahora observando el tercer término de (3.28) bajo la hipótesis (2) se tiene que $z_j = 0$ para todo $j \in I$, así que

$$\sum_{j \in I(x^*)} z_j u_j^k = 0 \quad (3.37)$$

finalmente, considerando la hipótesis (1) y se observa que $u^k \in \mathbb{R}^n$, $u^k \in \partial f(x^{k+1})$, y f es \mathcal{C}^1 así, $\partial f(x^{k+1}) = \{\nabla f(x^{k+1})\}$, por lo que $u^k = \nabla f(x^{k+1})$ de donde

$$u_j^k = \frac{\partial f(x^{k+1})}{\partial x_j^k},$$

así que $\lim_{k \rightarrow \infty} u_j^k = \frac{\partial f(x^*)}{\partial x_j}$.

Afirmación 1:

$$\frac{\partial f(x^*)}{\partial x_j} \geq 0 \quad j \in I(x^*)$$

En efecto; supóngase lo contrario es decir que $\frac{\partial f(x^*)}{\partial x_j} < 0$. Como $\frac{\partial f(x^*)}{\partial x_j} = \lim_{k \rightarrow \infty} u_j^k$ para un k suficientemente grande, se tiene $u_j^k < 0$, pero como $u_j^k = -\lambda_k (x_j^k)^{p-1} \varphi'(ct_j^k)$, se sigue que $\varphi'(ct_j^k) > 0$, lo que implica que $\varphi'(ct_j^k) > \varphi'(c)$ y como φ' es creciente, entonces $ct_j^k > c \Rightarrow t_j^k > 1$, así

$$x_j^{k+1} > x_j^k. \quad (3.38)$$

Finalmente se tiene que $\lim_{k \rightarrow \infty} x_j^k = x^* = 0$ para $j \in I(x^*)$ lo que contradice (3.30).

$$\lim_{k \rightarrow \infty} \sum_{j \in I(x^*)} z_j u_j^k = \sum_{j \in I(x^*)} z_j \frac{\partial f(x^*)}{\partial (x_j)} \geq 0. \quad (3.39)$$

Tomando el limite en (3.28) en vista a (3.31),(3.36),(3.37) y (3.39), se obtiene $f(x^*) - f(z) \leq 0$.

Puesto que $z \in T$, se concluye que $x^* \in T$. ■

3.4.2. Casí-Fejer convergencia.

La noción de Fejer convergencia de una sucesión $\{y^k\}$ en un conjunto V con respecto a la medida divergente d , viene dada por $d(v, y^{k+1}) \leq d(v, y^k)$ para todo $v \in V$ y toda $k \in \mathbb{N}$. El siguiente contra ejemplo demuestra que la sucesión $\{x^k\}$ definida por EPMG en el caso más sencillo con $p = c = k = 1$ no es Fejer convergente en el conjunto T con respecto a d_φ^p para cualquier $\varphi \in \Phi$

Ejemplo 3.4.1 (Ver Ejemplo 4.1, [11])

Sea $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definida como $f(x) = (x_1 - x_2)^2$ y tomando φ como sigue

$$\varphi(t) = \begin{cases} \frac{1}{16}(\frac{1}{16}t^3 - \frac{17}{16}t^2 + 7t - 6) & \text{si } t \in (0, \frac{1}{4}] \\ (t-1)^2 & \text{si } t \in (\frac{1}{4}, +\infty] \end{cases} \quad (3.40)$$

donde

$$\varphi'(t) = \begin{cases} \frac{1}{16}(\frac{3}{16}t^2 - \frac{34}{16}t + 7) & \text{si } t \in (0, \frac{1}{4}] \\ 2(t-1) & \text{si } t \in (\frac{1}{4}, +\infty] \end{cases} \quad (3.41)$$

y

$$\varphi''(t) = \begin{cases} \frac{1}{8}(\frac{3}{4}t - \frac{51}{8} + 14t^{-2}) & \text{si } t \in (0, \frac{1}{4}] \\ 2 & \text{si } t \in (\frac{1}{4}, +\infty] \end{cases} \quad (3.42)$$

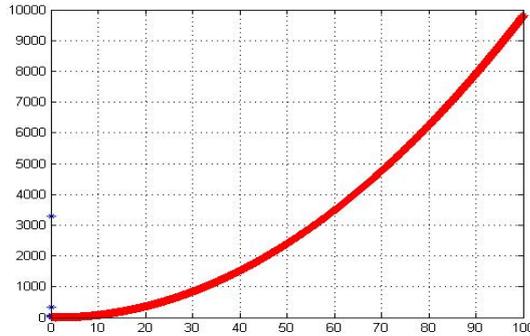


Figura 3.5: Contraejemplo de Fejér convergencia

En la figura 3.5 se observa que $\varphi \in \phi$.

Definición 3.4.1 (Ver Def. 4.1, [11])

Una sucesión $\{y^k\}$ es casi-Fejér convergente aun conjunto V con respecto a la medida divergente si para cada $v \in V$ existe $K \geq 0$ y una sucesión de números reales $\epsilon_k \geq 0$ tal que $\sum_{k=0}^{\infty} \epsilon_k < \infty$ y

$$d(v, y^{k+1}) \leq d(v, y^k) + \epsilon_k$$

para todo $k \geq K$

Teorema 3.4.1 (Ver Teo. 4.1, [11])

1. Si una sucesión $\{y^k\}$ es casi-Fejér convergente en un conjunto no-vacío V entonces $\{y^k\}$ es acotada.
2. Si V contiene un punto límite $\{y^k\}$ entonces $\{y^k\}$ converge

Demostración:

1. Para cualquier $v \in V$ y todo $i \geq K$

$$d(v, y^{k+1}) - d(v, y^k) \leq \epsilon_i \quad (3.43)$$

Sea $\delta = \sum_{k=K}^{\infty} \epsilon_k$. Tomando (3.77) con i entre K y k

$$d(v, y^{k+1}) \leq d(v, y^k) + \sum_{k=K}^{\infty} \epsilon_k \leq d(v, y^k) + \delta,$$

concluyendose por la definición de medida divergente que $\{y^k\}$ es acotada.

2. Sea y^* un punto limite de $\{y^k\}$ y $\{y^{l_k}\}$ una subsucesión de $\{y^k\}$ la cual converge a y^* . Dado $\epsilon > 0$ tomando $K_1 \geq K$ tal que $\sum_{k=K}^{\infty} \epsilon_k < \epsilon/2$ y \widehat{K} tal que $l_{\widehat{K}} \geq K_1$ y $d(y^*, y^{l_k}) \geq \epsilon/2$ para todo $k \geq \widehat{K}$ de la definición de medida divergente se puede asegurar la existencia de \widehat{K} . Puesto que $y^* \in V$ sustituyendo y^* por v e i entre l_k y k en (3.77) se tiene que

$$d(y^*, y^{k+1}) \leq d(y^*, y^{l_k}) + \sum_{i=l_k}^{\infty} \epsilon_i \leq \epsilon/2 + \epsilon/2 = \epsilon.$$

Puesto que ϵ es arbitrario, se tiene $\lim_{k \rightarrow \infty} d(y^*, y^k) = 0$, y de la definición de medida divergente inciso (2) esto es equivalente la cual es $\lim_{k \rightarrow \infty} y^k = y^*$

■

3.4.3. Resultados de Convergencia.

En esta sección se iniciará el estudio de los resultados de convergencia para *Una Casi-Distancia Generalizada en Programación Convexa. Caso $0 < p \leq 1$* , ya que las técnicas usada a partir de aquí para el caso $1 < p < 2$ no son recomendadas. Ver **Proposición**

3.4.5.

Obsérvese que el caso $p = c = k = 1$ fue estudiado en [11]. Vale la pena destacar que a partir de aquí se inicia el estudio para una familia de funciones la cual generaliza la función **Kullback-Leibler** (Ver el Ejemplo 3.3.1).

Sea $\gamma = \lim_{k \rightarrow \infty} f(x^k)$, el cual existe por el **Corolario 3.4.1** inciso (i).

Definiendo $T^* = \{x \in U : f(x) \leq \gamma\}$, claramente $T^* \supset T$ y T_1 como el conjunto de puntos límites de $\{x^k\}$. Por el **Corolario 3.4.1**(i) $T_1 \subset T^*$. Sea I el conjunto de índices j para la cual la sucesión $\{x_j^{j_k}\}$ de $\{x_j^k\}$ converge a 0. Sea $T_2 = \{x \in T^* : I \subset I(x)\}$, es decir, T_2 contiene a los puntos $x \in T^*$ tal que $\{x_j\} = 0$ para la componente j tal que $\{x_j^k\}$ no está acotada desde cero.

Proposición 3.4.5

Sea $\varphi \in \phi$ y denotando $\alpha := \frac{p}{c\varphi''(c)} > 0$ entonces existe un $\mu > 0$ tal que para todo $t \in \mathbb{R}_{++}$ se tiene que

$$t^p - 1 - \alpha t \varphi'(t) \leq \mu(\widehat{\varphi}(t) - \varphi(\bar{k})), \quad (3.44)$$

$$(t - t^p) + \alpha t \varphi'(t) - t \ln(t) \leq \mu(\widehat{\varphi}(t) - \varphi(\bar{k})). \quad (3.45)$$

Demostración:

Definiendo

$$\widetilde{\varphi}(t) = \frac{t^p - 1 - \alpha t \varphi'(t)}{c(t-1)\varphi'(ct)}, \quad (3.46)$$

$$\overline{\varphi}(t) = \frac{(t - t^p) + \alpha t \varphi'(t) - t \ln(t)}{c(t-1)\varphi'(ct)}. \quad (3.47)$$

Puesto que $\widetilde{\varphi}(t)$ y $\overline{\varphi}(t)$ son continuas para cualquier $t > 0$ excepto posiblemente $t = 1$, la existencia de μ esta garantizada si se demuestra que:

$\limsup_{t \rightarrow 0} \widetilde{\varphi}(t)$, $\lim_{t \rightarrow 1} \widetilde{\varphi}(t)$, $\lim_{t \rightarrow \infty} \widetilde{\varphi}(t)$ y $\limsup_{t \rightarrow 0} \overline{\varphi}(t)$, $\lim_{t \rightarrow 1} \overline{\varphi}(t)$, $\lim_{t \rightarrow \infty} \overline{\varphi}(t)$ son todos finitos.

Notese que φ' es estrictamente creciente, así que, $\limsup_{t \rightarrow 0} \varphi'(t) < 0$, $\limsup_{t \rightarrow \infty} \varphi'(t) > 0$ puesto que $\varphi'(\bar{k}) = 0$.

Reescribiendo (3.46)

$$\tilde{\varphi}(t) = \frac{t^p - 1}{c(t-1)\varphi'(ct)} - \frac{\alpha t}{c(t-1)}, \quad (3.48)$$

donde

$$\limsup_{t \rightarrow 0} \tilde{\varphi}(t) = \limsup_{t \rightarrow 0} \frac{1}{c\varphi'(ct)} < \infty,$$

por otra parte se tiene que;

$$\limsup_{t \rightarrow \infty} \tilde{\varphi}(t) = \limsup_{t \rightarrow \infty} \left(\frac{t^p - 1}{t-1} \frac{1}{c\varphi'(ct)} - \frac{\alpha t}{c(t-1)} \right),$$

para los valores de $0 < p < 1$ se tiene que;

$$\limsup_{t \rightarrow \infty} \frac{t^p - 1}{t-1} = 0, \quad \limsup_{t \rightarrow \infty} \frac{1}{c\varphi'(ct)} < \infty \quad y \quad \limsup_{t \rightarrow \infty} \frac{\alpha t}{c(t-1)} = -\frac{\alpha}{c},$$

por lo que

$$\limsup_{t \rightarrow 0} \tilde{\varphi}(t) = -\frac{\alpha}{c} < \infty,$$

por último nos falta estudiar,

$$\lim_{t \rightarrow 1} \frac{t^p - 1 - \alpha t \varphi'(ct)}{c(t-1)\varphi'(ct)},$$

este límite tiene forma indeterminada $0/0$, aplicando L'Hopital se, tiene que,

$$\lim_{t \rightarrow 1} \frac{pt^{p-1} - \alpha \varphi'(ct) - c\alpha \varphi''(ct)}{c\varphi'(ct) + c^2(t-1)\varphi''(ct)},$$

aplicando nuevamente L'Hopital, nos queda,

$$\begin{aligned} \lim_{t \rightarrow 1} \frac{p(p-1)t^{p-2} - c\alpha \varphi''(ct) - c\alpha \varphi''(ct) - c^2 t \alpha \varphi'''(ct)}{c^2 \varphi''(ct) + c^2 \varphi''(ct) + c^3(t-1)\varphi'''(ct)} \\ = \frac{p(p-1) - 2c\alpha \varphi''(c) - c^2 \alpha \varphi'''(c)}{2c^2 \varphi''(c)} \end{aligned}$$

sustituyendo el valor de α nos queda;

$$\lim_{t \rightarrow 1} \tilde{\varphi}(t) = \frac{\alpha}{2} \left(\frac{1}{c}(p-1) - \frac{1}{2c} - \frac{\varphi'''(c)}{\varphi''(c)} \right) < \infty.$$

De la desigualdad (3.45) se tiene que

$$(t - t^p) + \alpha t \varphi'(ct) - t \ln(t) \leq \mu(c(t-1)\varphi'(ct)),$$

definiendo,

$$\bar{\varphi}(t) = \frac{(t - t^p) + \alpha t \varphi'(ct) - t \ln(t)}{(c(t-1)\varphi'(ct))},$$

reescribiendo,

$$\bar{\varphi}(t) = \frac{(t - t^p)}{c(t-1)\varphi'(ct)} + \frac{\alpha t}{c(t-1)} - \frac{t \ln(t)}{c(t-1)\varphi'(ct)}.$$

Estudiamos el $\limsup_{t \rightarrow 0} \bar{\varphi}(t)$.

Obsérvese que para el primer sumando se tiene

$$\limsup_{t \rightarrow 0} \frac{(t - t^p)}{c(t-1)\varphi'(ct)} < 0,$$

para el segundo sumando

$$\limsup_{t \rightarrow 0} \frac{\alpha t}{c(t-1)} = 0,$$

y por último para el tercer sumando

$$\limsup_{t \rightarrow 0} -\frac{t \ln(t)}{c(t-1)\varphi'(ct)} = 0,$$

de acuerdo al convenio $0 \ln(0) = 0$ y el hecho que $\limsup_{t \rightarrow 0} \frac{1}{\varphi'(ct)} < 0$. Por lo que $\limsup_{t \rightarrow 0} \bar{\varphi}(t) < 0$.

Ahora se debe estudiar $\limsup_{t \rightarrow \infty} \bar{\varphi}$ esto es

$$\limsup_{t \rightarrow \infty} \frac{(t - t^p)}{c(t-1)\varphi'(ct)} + \frac{\alpha t}{c(t-1)} - \frac{t \ln(t)}{c(t-1)\varphi'(ct)}, \quad (3.49)$$

estudiando el primer sumando se tiene que para los valores de $0 < p < 1$,

$$\limsup_{t \rightarrow \infty} \frac{t - t^p}{t - 1} = 1, \quad \limsup_{t \rightarrow \infty} \frac{1}{c\varphi'(ct)} < \infty,$$

en el segundo sumando se tiene que

$$\limsup_{t \rightarrow \infty} \frac{\alpha t}{c(t-1)} = -\frac{\alpha}{c},$$

y por último para el tercer sumando se tiene que

$$\limsup_{t \rightarrow \infty} -\frac{t \ln(t)}{c(t-1)\varphi'(ct)} < \infty,$$

puesto que para $t > 1$, $\frac{t \ln(t)}{c(t-1)\varphi'(ct)} > 0$.

Concluyendo $\limsup_{t \rightarrow \infty} \bar{\varphi}(t) < \infty$.

Nos falta estudiar únicamente

$$\lim_{t \rightarrow 1} \frac{(t - t^p) + \alpha t \varphi'(ct) - t \ln(t)}{(c(t-1)\varphi'(ct))},$$

obsérvese que dicho límite posee la forma indeterminada $0/0$ y aplicando L'Hopital, se tiene que,

$$\lim_{t \rightarrow 1} \frac{pt^{p-1} - 1 - \alpha \varphi'(ct) + ct \alpha \varphi''(ct) - \ln(t) - 1}{c \varphi'(ct) + c^2(t-1)\varphi''(ct)},$$

para $\alpha = \frac{p}{c \varphi''(c)}$ y $0 < p < 1$ se tiene nuevamente la forma indeterminada $0/0$, así, que aplicando L'Hopital, se tiene que

$$\lim_{t \rightarrow 1} \frac{(p-1)t^{p-2} + c \alpha \varphi''(ct) + c \alpha \varphi''(ct) + c^2 t \alpha \varphi'''(ct) - (1/t)}{c^2 \varphi''(ct) + c^2 \varphi''(ct) + c^3(t-1)\varphi'''(ct)},$$

luego calculando el límite y sustituyendo el valor de α , se tiene que,

$$\lim_{t \rightarrow 1} \frac{(t - t^p) + \alpha t \varphi'(ct) - t \ln(t)}{(c(t-1)\varphi'(ct))} = \frac{\alpha}{c} - \frac{\alpha}{cp} + \frac{\alpha \varphi'''(c)}{2\varphi''(c)} < \infty,$$

Así $\limsup_{t \rightarrow 1} \bar{\varphi}(t) < \infty$. ■

Proposición 3.4.6 (Ver Prop. 4.4, [11])

Sea $z \in T_2$ cualquiera y μ como en la proposición (3.4.5). Entonces:

$$1. d_{\psi}^p(z, x^{k+1}) - d_{\psi}^p(z, x^k) \leq \frac{\mu}{\lambda} \lambda_k \sum_{j=1}^n \left(1 + \frac{z_j}{x_j^{k+1}}\right) x_j^k (\widehat{\varphi}(ct_j^k) - \varphi(\bar{k})), \quad \forall k \geq 0.$$

2. Si $\varphi'(ct) \leq \varphi''(c)\log(t) - (t^{p-1} - 1)$ para todo $t > 0$ entonces.

$$d_{\psi}^p(z, x^{k+1}) - d_{\psi}^p(z, x^k) \leq \frac{\mu}{\lambda} \lambda_k d_{\varphi}^p(x^{k+1}, x^k)$$

Demostración:

$$\begin{aligned} & d_{\psi}^p(z, x^{k+1}) - d_{\psi}^p(z, x^k) = \\ & \sum_{j=1}^n (x_j^{k+1})^{p-1} [z_j \ln(z_j) - z_j \ln(x_j^{k+1}) - z_j + x_j^{k+1}] - \sum_{j=1}^n (x_j^k)^{p-1} [z_j \ln(z_j) - z_j \ln(x_j^k) - z_j + x_j^k] = \\ & \sum_{j=1}^n (x_j^{k+1})^{p-1} z_j \ln(z_j) - (x_j^{k+1})^{p-1} z_j \ln(x_j^{k+1}) - (x_j^{k+1})^{p-1} z_j + (x_j^{k+1})^p - (x_j^k)^{p-1} z_j \ln(z_j) + \\ & (x_j^k)^{p-1} z_j \ln(x_j^k) + (x_j^k)^{p-1} z_j - (x_j^k)^p = \\ & \sum_{j=1}^n (x_j^{k+1})^{p-1} z_j \ln(z_j) - (x_j^{k+1})^{p-1} z_j \ln(x_j^{k+1}) - (x_j^{k+1})^{p-1} z_j + (x_j^{k+1})^p - (x_j^k)^{p-1} z_j \ln(z_j) + \\ & (x_j^k)^{p-1} z_j \ln(x_j^k) + (x_j^k)^{p-1} z_j - (x_j^k)^p + (x_j^k)^{p-1} z_j \ln(x_j^{k+1}) - (x_j^k)^{p-1} z_j \ln(x_j^{k+1}) = \\ & \sum_{j=1}^n (x_j^{k+1})^{p-1} [z_j \ln(z_j) - z_j - z_j \ln(x_j^{k+1})] - (x_j^k)^{p-1} [z_j \ln(z_j) - z_j - z_j \ln(x_j^{k+1})] - (x_j^k)^{p-1} z_j \ln\left(\frac{x_j^{k+1}}{x_j^k}\right) + \\ & (x_j^{k+1})^p - (x_j^k)^p \\ & \sum_{j=1}^n (x_j^k)^p \left[\frac{(x_j^{k+1})^{p-1}}{(x_j^k)^p} [z_j \ln(z_j) - z_j - z_j \ln(x_j^{k+1})] - (x_j^k)^{-1} [z_j \ln(z_j) - z_j - z_j \ln(x_j^{k+1})] - (x_j^k)^{-1} z_j \ln\left(\frac{x_j^{k+1}}{x_j^k}\right) + \right. \\ & \left. \frac{(x_j^{k+1})^p}{(x_j^k)^p} - 1 \right] = \\ & \sum_{j=1}^n (x_j^k)^p \left[\frac{z_j}{x_j^{k+1}} [((t_j^k)^p - t_j^k) \ln\left(\frac{z_j}{x_j^{k+1}}\right) - ((t_j^k)^p - t_j^k) - (t_j^k) \ln(t_j^k)] + t_j^p - 1 \right] \end{aligned} \quad (3.50)$$

Por otra parte, tomando u^k como lo definimos en (3.24). Por la **Proposición 3.2.1** y la definición de subgradiente se tiene;

$$(x^{k+1} - z)^T u^k \geq f(x^{k+1}) - f(z) \geq f(x^{k+1}) - \gamma \geq 0, \quad (3.51)$$

tomando $z \in T^*$ en la desigualdad anterior. Sea $\alpha = \frac{p}{c\varphi''(k)} > 0$. Multiplicando (3.51) por $\frac{\alpha}{\lambda_k}$ se obtiene,

$$0 \leq \frac{\alpha}{\lambda_k} (x^{k+1} - z)^T u^k = \frac{\alpha}{\lambda_k} \sum_{j=1}^n (x^{k+1} - z)^T u^k = \alpha \sum_{j=1}^n (x_j^k)^p \left(\frac{z_j}{x_j^{k+1}} - 1 \right) t_j^k \varphi'(ct_j^k), \quad (3.52)$$

de (3.50) y (3.52) se obtiene que

$$d_{\psi}^p(z, x^{k+1}) - d_{\psi}^p(z, x^k) \leq \sum_{j=1}^n (x_j^k)^p \left[\frac{z_j}{x_j^{k+1}} \left[((t_j^k)^p - t_j^k) \ln\left(\frac{z_j}{x_j^{k+1}}\right) - ((t_j^k)^p - t_j^k) - (t_j^k) \ln(t_j^k) + t_j^k \alpha \varphi'(ct_j^k) \right] + t_j^p - 1 - t_j^k \alpha \varphi'(ct_j^k) \right], \quad (3.53)$$

tomando a $z \in T_2$, la definición de T_2 implica que existe $\beta > 0$ tal que $x_j^k \geq \beta$ para todo $k \geq 0$, con $j \in J(z)$ Sea

$$\eta = \max_{1 \dots j \dots n} z_j, \\ \xi = \frac{1}{2} \min_{1 \dots j \dots n} z_j,$$

Por lo que,

$$\frac{\eta}{\beta} \geq \frac{z_j}{x_j^{k+1}} \geq \frac{\xi}{M}. \quad (3.54)$$

Considerese la siguiente sucesión $\{\alpha^k\}$ dada por

$$\alpha_j^k = \frac{(t_j^k - (t_j^k)^p) \ln\left(\frac{z_j}{x_j^{k+1}}\right) + t_j^k \alpha \varphi'(ct_j^k)}{t_j^k \varphi'(ct_j^k)} \quad \text{si } t_j^k \neq 1 \quad \text{para toda } k. \quad (3.55)$$

Afirmación 2:

La sucesión $\{\alpha^k\}$ es acotada además $\alpha^k > 0$ para toda k .

Prueba:

Observese en primer lugar que la sucesión $\{\alpha^k\}$ es acotada para ello estudiase el límite cuando $k \rightarrow \infty$ esto es,

$$\lim_{k \rightarrow \infty} \frac{(t_j^k - (t_j^k)^p) \ln\left(\frac{z_j}{x_j^{k+1}}\right) + t_j^k \alpha \varphi'(ct_j^k)}{t_j^k \varphi'(ct_j^k)}, \quad (3.56)$$

rescribiendo (3.56) se tiene que

$$\lim_{k \rightarrow \infty} \frac{(t_j^k - (t_j^k)^p) \ln\left(\frac{z_j}{x_j^{k+1}}\right)}{t_j^k} + \alpha. \quad (3.57)$$

Estudiando el primer múltiplo de (3.57)

$$\lim_{k \rightarrow \infty} \frac{(t_j^k - (t_j^k)^p)}{t_j^k} = 1, \quad (3.58)$$

del segundo término de (3.57)

$$\lim_{k \rightarrow \infty} \frac{\ln\left(\frac{z_j}{x_j^{k+1}}\right)}{\varphi'(ct_j^k)} = 0, \quad (3.59)$$

esto por (3.54) y puesto que $\lim_{k \rightarrow \infty} \varphi'(ct_j^k) = +\infty$. Así de (3.57) y (3.58) se tiene que $\{\alpha^k\}$ es acotada.

Nos falta probar que $\alpha^k > 0$ para toda $k \in \mathbb{N}$ basta observar que para un K_0 suficientemente grande se tiene que $\alpha_j^k > 0$; lo que prueba la **Afirmación 2**.

Así para un $\alpha_j^k > 0$ tal que $\alpha = \frac{p}{c\varphi''(c)} < \alpha_j^k$ de (3.53) se tiene

$$d_{\psi}^p(z, x^{k+1}) - d_{\psi}^p(z, x^k) \leq \sum_{j=1}^n (x_j^k)^p \left[\frac{z_j}{x_j^{k+1}} [-(t_j^k)^p + t_j^k - (t_j^k) \ln(t_j^k) + t_j^k \alpha_j^k \varphi'(ct_j^k)] + t_k^p - 1 - t_j^k \alpha \varphi'(ct_j^k) \right]. \quad (3.60)$$

Usando las desigualdades (3.44) y (3.45) en (3.60) se tiene que

$$d_{\psi}^p(z, x^{k+1}) - d_{\psi}^p(z, x^k) \leq \frac{\mu}{\lambda} \lambda_k \sum_{j=1}^n \left(1 + \frac{z_j}{x_j^{k+1}}\right) (x_j^k)^p (\widehat{\varphi}(ct_j^k) - \varphi(\bar{k})), \quad \text{para toda } k \geq 0.$$

hemos probado el inciso 1. Para probar 2, usando la hipótesis $\varphi'(ct) \leq \varphi''(c) \log(t) - (t^{p-1} - 1)$ y (3.60) se tiene

$$d_{\psi}^p(z, x^{k+1}) - d_{\psi}^p(z, x^k) \leq \mu c^2 \sum_{j=1}^n (x_j^k)^{p-1} (x_j^{k+1} - x_j^k) \varphi'(ct_j^k) \leq c^2 \frac{\mu}{\lambda} \lambda_k d_{\psi}^p(x^{k+1}, x^k).$$

■

En el orden de establecer cuasí-Fejér convergencia a T_1 se darán resultados preliminares los cuales serán usados para establecer la convergencia local.

Para cualquier $z \in T^*$. Define:

$$\xi = \frac{1}{2} \min_{j \in J(z)} z_j, \quad (3.61)$$

$$\eta = \max_{1 \leq j \leq n} z_j, \quad (3.62)$$

$$\tilde{\mu} = \frac{\mu}{\lambda} \left(1 + \frac{2\eta}{\xi}\right), \quad (3.63)$$

con μ como en la **Proposición 3.4.5**. Tomando $\sigma > 0$ tal que $d_\varphi^p(z, x) \leq \sigma$ implica $x_j \geq \xi$ para todo $j \in J(z)$.

Proposición 3.4.7 (Ver Prop.4.5, [11])

Si existe K_0 tal que:

$$f(x^{K_0}) - \gamma \leq \min\left\{\frac{\sigma}{2\tilde{\mu}}, \hat{\lambda} \frac{\xi}{c} (\varphi(\frac{c}{2}) - \varphi(c))\right\} \quad y \quad (3.64)$$

$$d_\varphi^p(z, x) \leq \frac{\sigma}{2}, \quad (3.65)$$

y μ como en la **Proposición (3.4.5)**. Entonces: para todo $k \geq K_0$,

$$d_\varphi^p(z, x^{k+1}) - d_\varphi^p(z, x^k) \leq \tilde{\mu} \lambda_k d_\varphi^p(x^{k+1}, x^k) \quad (3.66)$$

Demostración:

Primero se afirmara que (3.66) se tiene para $k \geq K_0$ tal que

$$d_\varphi^p(z, x^k) \leq \sigma. \quad (3.67)$$

Esto será probado más adelante. Por definición de σ , para k tal que (3.67) se tiene:

$$x_j^k \geq \xi \quad (j \in J(z)), \quad (3.68)$$

entonces usando (3.68), la **Proposición 3.4.2**, el **Corolario 3.4.1** (1), (3.64) y ($k \geq K_0$) para $j \in J(z)$

$$\begin{aligned} \frac{\xi^p}{c} (\varphi(ct_j^k) - \varphi(c)) &\leq \frac{(x_j^k)^p}{c} (\varphi(ct_j^k) - \varphi(c)) \leq \\ \frac{\lambda_k}{\lambda} d_\varphi^p(x^{k+1}, x^k) &\leq \frac{\lambda_k}{\lambda} d_\varphi^p(x^{k+1}, x^k) \leq \frac{1}{\lambda} (f(x^k) - f(x^{k+1})) \leq \\ \frac{1}{\lambda} (f(x^{K_0}) - \gamma) &\leq \frac{\xi^p}{c} (\varphi(\frac{c}{2}) - \varphi(c)), \end{aligned} \quad (3.69)$$

de (3.69) se tiene, $(\varphi(ct_j^k) - \varphi(c)) \leq (\varphi(\frac{c}{2}) - \varphi(c))$, o lo que es igual a, $\varphi(ct_j^k) \leq \varphi(\frac{c}{2})$. Puesto que φ es decreciente en $(0, c)$ se concluye que $t_j^k \leq \frac{1}{2}$ para $j \in J(z)$. De (3.68) se tiene $x_j^{k+1} = x_j^k t_j^k \geq \frac{\xi}{2}$ para $j \in J(z)$ y ademas usando (3.62) y el hecho de que $\{x_j^k\}$ es acotada se tiene que,

$$\frac{z_j}{x_j^{k+1}} \leq \frac{2\eta}{\xi}, \tag{3.70}$$

para todo j .

Usando las desigualdades (3.60) y la sucesión dada en (3.55), para todo $k \geq K_0$ se tiene que

$$t_j^k \ln\left(\frac{z_j}{x_j^{k+1}}\right) - (t_j^k)^p \ln\left(\frac{z_j}{x_j^{k+1}}\right) > t_j^k \ln\left(\frac{2\eta}{\xi}\right) - (t_j^k)^p \ln\left(\frac{\xi}{M}\right) \tag{3.71}$$

luego usando (3.63) nosotros obtenemos (3.66).

Falta verificar que (3.66) se tiene para todo $k \geq K_0$. Procedamos por inducción, (3.65) implica $d_\varphi^p(z, x^{K_0}) \leq \sigma$ por lo que (3.68) es cierto para $k = K_0$. Supongase que (3.68) se cumple para k entre K_0 e i . Y pruébese que (3.68) se cumple para $k = i + 1$ así;

$$d_\varphi^p(z, x^{i+1}) - d_\varphi^p(z, x^{K_0}) \leq \tilde{\mu} \sum_{k=K_0}^i \lambda_k d_\varphi^p(x^{k+1}, x_k),$$

$$\tilde{\mu}(f(x^{K_0}) - f(x^{i+1})) \leq \tilde{\mu}(f(x^{K_0}) - \gamma) \leq \frac{\sigma}{2},$$

usando la **Proposición 3.4.2**, el **Corolario 3.4.1**(1), (3.64). De (3.70) y usando (3.65):

$$d_\varphi^p(z, x^{i+1}) \leq d_\varphi^p(z, x^{K_0}) + \frac{\sigma}{2} = \frac{\sigma}{2} + \frac{\sigma}{2} = \sigma \tag{3.72}$$

(3.72) demuestran que (3.66) se cumple para $k = i + 1$ y de igual forma se cumple (3.66). ■

Observar que de las dos proposiciones anteriores se obtiene la misma cota superior para $z \in T_1$ y $z \in T_2$, los siguientes corolarios son consecuencia de estas proposiciones estableciendo la casi-Fejer convergencia en T^* con respecto a d_ψ^p

Corolario 3.4.2 (Ver Coro. 4.2, [11])

1. $\{x^k\}$ es casi-Fejer convergente en T_2 con respecto a d_ψ^p
2. Si $\varphi'(ct) \leq \varphi''(c)\log(t) - (t^{p-1} - 1)$ para todo $t > 0$ entonces $\{x^k\}$ es cuasi-fejer convergente a T^* con respecto a d_ψ^p

Demostración:

Aplicando la **Proposición 3.4.6** (i) con $z \in T_2$. Considerese $\beta > 0$ tal que $\{x_j^k\}$ para todo $k \geq 0$, $j \in J(z)$, entonces $\frac{z_j}{x_j^{k+1}} \leq \frac{z_j}{\beta}$ para todo j y todo $k \geq 0$. Sea $\eta = \max_{1 \dots j \dots n} z_j$, así que para todo $k \geq 0$ y todo j

$$\frac{z_j}{x_j^{k+1}} \leq \frac{\eta}{\beta}, \quad (3.73)$$

de (3.73) y la **Proposición 3.4.6**(i) se tiene

$$d_\psi^p(z, x^{k+1}) - d_\psi^p(z, x^k) \leq \frac{\mu}{\lambda} \lambda_k \sum_{j=1}^n \frac{x_j^k}{c} \left[\left(1 + \frac{\eta}{\beta}\right) (\widehat{\varphi}(ct) - \varphi(\bar{k})) \right], \quad (3.74)$$

$$d_\psi^p(z, x^{k+1}) - d_\psi^p(z, x^k) \leq \sum_{j=1}^n x_j^k (t_j^k - 1) \widehat{\varphi}'(ct) = \widehat{\mu} \lambda_k d_{\widehat{\varphi}}^p(x^{k+1}, x^k), \quad (3.75)$$

con $\widehat{\mu} = \frac{\mu}{\lambda} \lambda_k (1 + \frac{\eta}{\beta})$. La definición de cuasi-Fejér convergencia se satisface con $K = 0$ y $\epsilon_k = \widehat{\mu} \lambda_k d_{\widehat{\varphi}}^p(x^{k+1}, x^k)$. El inciso dos se sigue directamente de la **Proposición 3.4.6** y el

Corolario 3.4.1. ■

Corolario 3.4.3 (Ver Coro. 4.3, [11])

1. $\{x^k\}$ es casi-Fejér convergente a T_1 con respecto a d_φ^p .
2. Si x^0 esta suficientemente cerca de el punto $z^* \in T$ entonces $\{x^k\}$ es casi-Fejér convergente a z^* con respecto a d_φ^p

Demostración:

1. Tomese $z \in T_1$, así existe una subsucesión $\{x^{l_k}\}$ de $\{x^k\}$ la cual converge a z . Por la definición de γ , la **Proposición 3.2.1** y la definición de medida divergente, existe \widehat{k} tal que (3.64), (3.65) se tiene para $K_0 = l_{\widehat{k}}$. La conclusión se sigue de (3.66), el **Corolario 3.4.1(1)** y la definición de casi-fejer convergente con $K = l_{\widehat{k}}$ y $\epsilon_k = \widetilde{\mu} \lambda_k d_{\varphi}^p(x_j^{k+1}, x_j^k)$.
2. Sea W una vecindad de z^* tal que para todo $x \in W$

$$f(x) - f(z^*)\gamma \leq \min\left\{\frac{\sigma}{2\widetilde{\mu}}, \widehat{\lambda} \frac{\xi}{c} (\varphi(\frac{c}{2}) - \varphi(c))\right\}, \quad (3.76)$$

$$d_{\varphi}^p(z^*, x) \leq \frac{\sigma}{2}, \quad (3.77)$$

(3.76) puede ser logrado por la continuidad de f en z^* y (3.77), **Proposición 3.2.1** y la definición de medida divergente. Así si $x^0 \in W$, (3.64) y (3.65) se tienen con $K_0 = 0$. La conclusión se sigue como en el caso (1). ■

Teorema 3.4.2 (Ver Teo. 4.2, [11])

Si cualquiera de las siguientes condiciones se tiene:

1. *Existe una solución z del problema (D) tal que $\{x_j^k\}$ es acotada lejos de cero para componentes j tal que $z_j > 0$*
2. *$\varphi'(ct) \leq \varphi''(c)\log(t) - (t^{p-1} - 1)$ para todo $t > 0$*
3. *x^0 esta suficientemente cerca de el punto z^* del problema (P).*

entonces la sucesión $\{x^k\}$ generada por (3.12) y (3.13) convergen. Más aún, bajo cualquiera de estas condiciones, o si f continuamente diferenciable en \mathbb{R}^n , el límite de la sucesión es una solución del problema (D).

Demostración:

La condición (1) implica que T_2 es no vacío y así bajo (1), (2) y (3) se tiene la casi-Fejer

convergencia en el conjunto T_2 , T^* y $\{z^*\}$ con respecto a d_ψ^p por el **Corolario 3.4.2(1)**, **Corolario 3.4.2(2)** y el **Corolario 3.4.3(2)** respectivamente.

Debido a que $\{x_j^k\}$ es acotada por lo que tiene un punto límite así T_1 es no vacía y por el **Corolario 3.4.3(1)** se tiene la casi-fejer convergencia en el conjunto de puntos límites con respecto a d_ψ^p .

La siguiente afirmación que se hará es que la condición (2) y (3) implica la condición (1). Nuevamente por el **Colorario 3.4.2(2)**, **Corolario 3.4.3(2)** se obtiene la casi-Fejér convergencia en T^* y $\{z^*\}$ bajo (2) y (3) respectivamente. Así usando el **Teorema 3.4.1(1)** se obtiene que $\{d_\varphi^p(z, x^*)\}$ es acotado para todo $z \in T \subset T^*$ bajo (2) y $\{d_\varphi^p(z^*, x^*)\}$ es acotado bajo (3). Así pues, cuando (2) se tiene se cumple (1) para cualquier $z \in T$, y cuando (3) se tiene se cumple (1) para $z = z^*$. Lo próximo en notar es que si $x^* = \lim_{k \rightarrow \infty} x^k$, la condición (1) implica que $I(x^*) \subset I(z)$, esto se sigue de la **Proposición 3.4.4(1)** con $x^* \in T$ bajo (2),(3) o (5). Si f es continua diferenciable entonces el resultado se sigue inmediatamente de la **Proposición ??(2)**. ■

Corolario 3.4.4 (Ver Coro. 4.4, [11])

Si $\varphi'(ct) \leq \varphi''(c)\log(t) - (t^{p-1} - 1)$ para todo $t > 0$ entonces la sucesión generada por (1.12) y (1.13) converge a la solución de el problema (D).

Demostración:

La prueba es consecuencia inmediata del **Teorema ??**. ■

3.5. Convergencia desde el punto de vista Primal

Considere el problema (P) con hipótesis (H1) y (H2) y la θ – funciones cumpliendo las condiciones de la función de penalidad establecida en [19] en la sección 3.3.3. Dando $p \geq 0$, $r \in (0, 1]$, $\tilde{y} = 0$, $c \in \mathbb{R}_{++}$ con $\theta'(\tilde{y}) = c = k$, la función Langrageana aumentada está dada por

$$x^0 \in \mathbb{R}^n \quad (3.78)$$

$$L_{r,c}(x, \mu) = f_0(x) + \sum_{i=1}^m P_{p,i}(f_i(x), \mu_i, r, c) \quad (3.79)$$

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

El método de multiplicadores asociado al método de punto proximal está dado por:

$$x^{k+1} \in \operatorname{argmin}\{f_0(x) + \sum_{i=1}^m P_{\mu,i}(f_i(x), \mu_i^k, r^k, c)\} \quad (3.80)$$

$$\mu_i^{k+1} = \frac{\mu_i^k}{c} \theta' \left(\frac{f_i(x^{k+1})}{(\mu_i^k)^{p-1} r^k} + \tilde{y} \right) \quad \text{para } i = 1, \dots, m \quad \text{y } \mathbb{R}^k \in (0, 1]. \quad (3.81)$$

Observar que

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

donde l es la función Lagrangeana.

El siguiente resultado es tomado de [19] **Teorema (3.3.3.1)** el cual establece que la sucesión generada en (3.79) y (3.80) es la misma a la obtenida en el algoritmo de EPMG es decir;

$$\mu^0 \in \mathbb{R}_{++}^n \quad (3.82)$$

$$\mu^{k+1} = \arg \max_{\mu \geq 0} \{c(\mu) - r_k d_\varphi^p(\mu, \mu^k)\} \quad (3.83)$$

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

Proposición 3.5.1 (Ver Prop. (3.3.3.1), [19])

Sea $\{\hat{\mu}^k\}$ una sucesión generada por (3.83) que resuelve el problema dual (D) y sean $\{x^k\}$ y $\{\mu^k\}$ las sucesiones generadas por (3.80) y (3.81) que resuelve el problema primal (P).

Si $\mu^0 = \hat{\mu}^0$, entonces $\forall k > 0 : \mu^k = \hat{\mu}^k$

Demostración: Ver Prop. (3.3.3.1), [19].

Combinando la **Proposición 3.5.1**, con el **Teorema 3.4.2** podemos probar la convergencia de el método (3.80)-(3.81).

Teorema 3.5.1 Considerando el problema convexo (3.79). Sea $\varphi \in \Phi$ y supóngase que cualquier condición del **Teorema 3.4.2** es satisfecha para el problema (D). Entonces la sucesión μ^k converge a una solución óptima dual de μ^* , y si f es continua, la $\{x^k\}$ converge a la solución del problema (P).

Demostración:

De la **Proposición 3.4.8** se tiene que los multiplicadores óptimos son obtenidos aplicando el **EPMG** al problema dual de (P).

Por la hipótesis (H2) se tiene que el conjunto de multiplicadores lagrangiano es no vacío y compacto, así podemos aplicar el **Teorema 3.4.1(ii)** se obtiene la convergencia de $\{\mu^k\}$.

Más aun, si cualquiera de las condiciones del **Teorema 3.4.2** se tiene para (D) obtenemos la convergencia de $\{\mu^k\}$ a la solución óptima $\{\mu^*\}$ de (D).

Por otra parte sea \bar{x} el punto límite de $\{x^k\}$. Note que (3.79) es equivalente a

$$\mu^{k+1} = \arg \max_{\mu > 0} \{ \langle \mu, f(x^{k+1}) \rangle - \lambda_k d_{\varphi}^p(\mu, \mu^k) \}, \quad (3.84)$$

Escribiendo las condiciones de KKT para (3.84) se obtiene para cada $j = 1, \dots, n$

$$f_j(x^{k+1}) - \lambda_k (\mu_j^k)^{p-1} \varphi' \left(\frac{c \mu_j^{k+1}}{\mu_j^k} \right) + \eta_k = 0, \quad (3.85)$$

$$\eta_k \mu_{k+1} = 0, \quad (3.86)$$

$$\mu_j^{k+1} \geq 0, \quad (3.87)$$

$$\eta_k \geq 0. \quad (3.88)$$

Puesto que $\mu^k \geq 0$ para todo k , por la **Proposición 3.4.8** y **Proposición 3.4.1** se obtiene de (3.85) y (3.86)

$$0 = f_j(x^{k+1}) - \lambda_k(\mu_j^k)^{p-1} \varphi' \left(\frac{c\mu_j^{k+1}}{\mu_j^k} \right) = f_j(x^{x+1}) + u_j^k, \quad (3.89)$$

con u_j^k como la definimos en (3.24). Como μ^k converge a μ^* , nosotros podemos tomar una subsucesión x^{l_k} de x^k la cual converge a \bar{x} . Sea $\mu_j^k \rightarrow \mu_j^* > 0$ usando (3.85)-(3.87), $\varphi'(c) = 0$ y la continuidad de f se obtiene que $f(\bar{x}) \leq 0$.

Considérese ahora el caso donde para algún j , $\mu_j^k \rightarrow \mu_j^* = 0$. Entonces para este j , sea una sucesión $\mu_j^{l_k}$ tal que $\mu_j^{l_{k+1}} < \mu_j^{l_k}$ se tiene que $\varphi' \left(\frac{c\mu_j^{l_{k+1}}}{\mu_j^{l_k}} \right) < 0$, (ya que $\varphi'(c) = 0$ y φ' es creciente). Por lo que (3.80) donde $\tilde{y} = 0$ se obtiene que

$$\frac{c\mu_j^{k+1}}{\mu_j^k} = \theta' \left(\frac{f_j(x^{k+1})}{(\mu_j^k)^{p-1} r^k} \right), \quad (3.90)$$

$$(\theta')^{-1} \left(\frac{c\mu_j^{k+1}}{\mu_j^k} \right) = \left(\frac{f_j(x^{k+1})}{(\mu_j^k)^{p-1} r^k} \right), \quad (3.91)$$

luego por el **Corolario 23.5.1** en [17] se tiene que

$$f_j(x^{k+1}) = (\mu_j^k)^{p-1} r^k \varphi' \left(\frac{c\mu_j^{k+1}}{\mu_j^k} \right), \quad (3.92)$$

se obtiene que $f_j(x^{j_k}) < 0$. Ya se ha probado que \bar{x} es factible. Para completar la prueba se demostrara que \bar{x} es un minimizador de Lagrange. La condición de optimalidad de (3.76) en x^{k+1} es equivalente a

$$0 \in \partial \left\{ f(x^{k+1}) + \sum_{i=1}^m P_{\mu_i}(f_i(x^{k+1}), \mu_i^k, r^k, c) \right\}, \quad (3.93)$$

aplicando calculo de subgradiente y usando (3.72), se sigue de (3.93).

$$0 \in \partial \left\{ f(x^{k+1}) + \sum_{i=1}^m \frac{(\mu_i^{k+1})^p}{c} \left(\theta \left(\frac{f_i(x^{k+1})}{(\mu_i^{k+1})^{p-1} r} \right) - \theta(\tilde{y}) \right) \right\}, \quad (3.94)$$

$$0 \in \partial \left\{ f(x^{k+1}) + \sum_{i=1}^m \frac{(\mu_i^{k+1})^p}{c} \left(\varphi^* \left(\frac{f_i(x^{k+1})}{(\mu_i^{k+1})^{p-1} r} \right) - \varphi^*(\tilde{y}) \right) \right\}, \quad (3.95)$$

$$0 \in \partial f(x^{k+1}) + \sum_{i=1}^m \frac{(\mu_i^{k+1})^p}{c} ((\varphi^*)' \left(\frac{f_i(x^{k+1})}{(\mu_i^{k+1})^{p-1} r} \right) \frac{1}{\mu_i^{k+1})^{p-1} r} \partial f(x^{k+1})), \quad (3.96)$$

$$0 \in \partial f(x^{k+1}) + \sum_{i=1}^m \frac{(\mu_i^{k+1})^p}{c} ((\varphi')^{-1} \left(\frac{f_i(x^{k+1})}{(\mu_i^{k+1})^{p-1} r} \right) \frac{1}{\mu_i^{k+1})^{p-1} r} \partial f(x^{k+1})), \quad (3.97)$$

$$0 \in \partial f(x^{k+1}) + \sum_{i=1}^m (\mu_i^{k+1})^p \partial f(x^{k+1}) \quad (3.98)$$

tomando el límite de la subsucesión y usando la semicontinuidad del subdiferencial, obtenemos que:

$$0 \in \partial f(x^{k+1}) + \sum_{i=1}^m (\mu_i^*)^p \partial f(x^{\bar{x}}), \quad (3.99)$$

demostrando que \bar{x} minimiza a $L(x, \mu^*)$

■

Capítulo 4

Conclusiones.

A lo largo de esta investigación la primera conclusión obtenida es la estrecha relación que guardan entre si las casi-distancias asociadas a φ -divergencias, a núcleos homogéneos de segundo orden y a distancias de Bregman, ya que son casos particulares de una casi-distancia generalizada desarrollada en [19]. Además tal casi-distancia es obtenida de una familia de funciones que no necesariamente pasan por el origen ni necesariamente tienen pendiente uno. En [19] se obtiene la convergencia para el caso $p \geq 2$

En estudio de nuevos caminos para la convergencia de las casi-distancias generalizadas, se obtuvo la convergencia para valores de p que permanecían abiertos, ver [19], caso ($0 < p \leq 1$) y a través de una familia de funciones de penalidad que necesariamente no pasa por el origen, generalizando los resultados obtenidos en [11]. Sin embargo, las técnicas utilizadas en [11] no resultan convenientes para obtener resultados de convergencia para todo $c > 0$, por problemas de no acotabilidad, pero en el caso particular de $c = \bar{k}$ se obtiene $\tilde{y} = 0$ lo que equivale a una traslación vertical y obteniendo resultados para $0 < p \leq 1$ lo cual es más amplio que [11] donde se mostró solamente el caso $p = 1$ y , se deja así nuevos trabajos abiertos para el caso $1 < p < 2$, en el cual se sugiere estudiar nuevas técnicas para obtener dicha convergencia, igualmente estudiar nuevas vías, para un shift constante diferente al mínimo de la función de penalidad trasladada.

Por otro lado el método de punto proximal desarrollado con la casi-distancia generalizada sugiere que se debe estudiar métodos de punto proximal asociados a las familias de distancias de Bregman que dependan de un parámetro, como se observó que ocurre con la casi-distancia generalizada.

Bibliografía

- [1] A. Auslender, M. Teboulle and S. Ben-Tiba. *Interior proximal and multiplier methods based on second order homogeneous kernels*. Mathematics of Operations Research 24, No. 3, pp. 645-669, 1999.
- [2] M. S. Bazaraa and C.M. Shetty, *Nonlinear Programming, Theory and algorithms*. Georgia Institute of technology, Atlanta,1979.
- [3] A. M. Ben Tal and Zibulevsky. *Penalty / barrier multiplier methods for convex programming problems*. SIAM Journal on Optimization 7, pp. 447-336, 1997.
- [4] D.P. Bertsekas. *Constrained Optimization and Lagrange multiplier Methods*. Academic Press, New York, 1982.
- [5] J. Campos, R.A. Castillo and E. Hernández. *A multiplier method with a third order homogeneous kernel*. Technical Report, U.C.L.A., Venezuela, 2005.
- [6] R.A. Castillo and C.Gonzaga. *Penalidades generalizadas e métodos de lagrangeano aumentado para promamacao nao linear*. Dsc. these, U.F.R.J., Brasil, 1998.
- [7] J. Eckstein. *Nonlinear proximal point algoritms using Bregman functions, with applications to convex programming*. Mathematics of Operations REsearchs,18, pp. 202-226, 1993.
- [8] A. Fiacco and G. McCormick. *Nonlinear programmim: Sequential Unconstrained Minimization Techniques*. Classics in applied Mathematics, SIAM, Philadhelphia, 1990.

- [9] A. Iusem. *Metodos do ponto proximal em Otimizacao*. 20 Coloquio Brasileiro de Matemática, IMPA, R.J., Brazil, 1995.
- [10] K. Kiwiel. *Proximal minimizations methods with generalized Bregman functions*. SIAM J. on Control and optimization, 35, pp. 1142-1168, 1997.
- [11] A. Iusem, M. Teboulle and B. Svaiter. *Entropy-like proximal methods in convex programming*. Mathematics of Operations Research, 19(4), pp. 790-814, 1994.
- [12] A. Iusem. *Augmented Lagrangian methods and proximal point methods for convex optimization*. Investigación Operativa, 8, pp. 11-50, 1997.
- [13] Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithm II*. Springer Verlag, New York, 1996.
- [14] L. Matioli. *Tese de Doutorado, Universidade Federal de Santa Catarina*. Brazil, 2000.
- [15] R. Poliak and M. Teboulle. *Nonlinear rescaling and proximal-like methods in convex programming*. 76, pp. 667-739, 1997.
- [16] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, N.Y., 1970.
- [17] P.J. Silva, J. Eckstein and C. Humes. *Rescaling and Step size selection in proximal methods using separable generalized distances*. SIAM Journal on Optimization 12(1), pp. 238-261, 2001.
- [18] M. Teboulle. *Entropic proximal mappings with applied to nonlinear programming*. Mathematics of Operations Research 17, pp. 97-116, 1992.
- [19] R. A. Castillo, E. Hernández and J. Campos. *A Generalized Like-Distance in Convex Programming*. Internacional Mathematical Forum, 2, 2007, No. 37, pp. 1811-1830.
- [20] Moreau J. *Proximité et dualité dans un espace Hilbertien*. Bull. Soc. Math. France 93 pp 97-116, 1.976.

- [21] Martinet B. *Relaxation of variational inequalities by successive approximations*. Revue Française de Informatique et Recherche Oerationelle 2, pp. 154-159, 1.970.
- [22] R. T. Rockafellar. *Monotone operators and the proximal point algorithm in convex programming*. SIAM J. on Control and Optimization 14, pp. 877-898, 1.976.