Universidad Central de Venezuela
Facultad de Ciencias
Postgrado en Matemática

Una Casi-Distancia Generalizada en Programación Convexa

Autor: Eibar R. Hernández R.

Tutor: Rómulo A. Castillo C.

Tesis Doctoral

Presentada ante la Ilustre

Universidad Central de Venezuela

Para optar al título de

Doctor en Ciencias

Mención Matemática

Caracas, 07/12/2007

Universidad Central de Venezuela
Facultad de Ciencias
Postgrado en Matemática

Una Casi-Distancia Generalizada en Programación Convexa

Autor: Eibar R. Hernández R.

Tutor: Rómulo A. Castillo C.

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 para conseguir esta casi-distancia proviene de estudiar funciones de penalidad con shift constante en el espacio primal. Estas funciones de penalidad no necesariamente tienen que pasar a través del origen con pendiente uno y su función conjugada toma valores negativos. Para un caso particular se consigue una generalización de la distancia entrópica Kullback-Leibler. Esta casi-distancia generalizada puede ser vista en cada iteración del algoritmo, como la diferencia entre una función de Bregman y su aproximación lineal. Se muestran resultados de convergencia primal y dual, particularmente se muestra que cada punto límite de la sucesión generada por el método proximal definido por la casi-distancia generalizada es una solución óptima dual.

Caracas, 07/12/07

Dedicado a mi esposa Katiuska Massiel y a mis hijos
Eibar Daniel y Eibar Jesús

Agradecimiento

Mi agradecimiento primeramente a Dios que hace posible cada momento en mi vida y cada logro en mi carrera, a mis padres Eibar Timoshenko y Carmen Libertad por su cercanía y apoyo constante, a mi esposa Katiuska Massiel por su inmensa comprensión y por haber dedicado una gran cantidad de horas de su tiempo en la transcripción de este trabajo, a mi hijo Eibar Daniel, a quien quise inspirar en el estudio y razón escencial por la que inicié mis estudios de Doctorado, a mi hijo Eibar Jesús, por llenar de alegría mi vida con sus ocurrencias, al Dr. Rómulo Castillo, por su colaboración, disposición, dedicación y guía en el desarrollo de este trabajo y finalmente a todos aquellos que de una u otra forma colaboraron en la realización de este trabajo.

Introducción

En el marco de la programación convexa existe una estrecha relación entre los métodos de Lagrangeano aumentado y los métodos de punto proximal vía dualidad de Fenchel.

Por un lado, los métodos de Lagrangeano aumentado se han desarrollado por medio de diversas funciones de penalidad, por otro, los métodos de punto proximal se han desarrollado por medio de casi-distancias, asociadas a φ-divergencias, a núcleos homogéneos de sengundo orden y distancias de Bregman. Estas casi-distancias son utlizadas, entre otras cosas, para resolver el problemal dual en dicha programación.

El problema primal con desigualdades en programación convexa viene dado por:

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

donde $f_i : \mathbb{R}^n \to \mathbb{R}$, para $i = 0, 1, \ldots, m$ son funciones convexas, propias y cerradas.

El problema dual asociado al problema primal viene dado por:

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

donde $d(\mu) = \inf\{l(x, \mu) : x \in \mathbb{R}^n\}$ y $l(x, \mu)$ es la función Lagrangeana, la cual esta dada por:

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

Los problemas primal y dual han sido estudiados utilizando básicamente dos puntos de vista, el punto de vista primal-dual, [3],[4],[16] y el punto devista dual-primal, [1],[7],[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 actuan como núcleos. Como se expresó anteriormente, estas casi-distancias están asociadas a φ-divergencias, núcleos cuadráticos y distancias de Bregman y es de hacer notar que actualmente guardan poca relación entre si y los estudios de convergencia en cada caso son realizados por vías muy diferentes, de ahí que surja la inquietud de deterninar si tales casi-distancias representan objetos matemáticos totalmente distintos, o por el contrario guardan una estrecha relación entre ellas. Sorprendentemente lo que ocurre es aún mejor. Todas ellas son casos particulares de una casi-distancia general.

En las siguientes páginas se encontrará la teoría para desarrollar una casi-distancia generalizada que contiene como casos particulares las mencionadas anteriormente, la cual proviene de estudiar funciones de penalidad, que no necesariamente pasan a través del origen con pendiente 1, por medio de un shift constante en el espacio primal al aplicar la teoría de dualidad de Fenchel.

En el Capítulo 1 se desarrollarán algunos conceptos y teoremas básicos del análisis convexo y la programación no lineal, suficientes para ofrecer una mejor comprensión de esta investigación.

En el Capítulo 2 se estudiará la relación entre los métodos de Lagrangeano aumentado y los métodos de punto proximal, ofreciendo también en este capítulo una breve exposición del desarrollo histórico de ambos.

Finalmente, en el Capítulo 3 se presentan los aportes originales de este trabajo, ofreciendo una familia de funciones de penalidad para el método de Lagrangeano aumentado y se construye la casi-distancia generalizada vía dualidad de Fenchel, incluyendo un método y los teoremas de convergencia en los espacios dual y primal.

Índice general

Índice de figuras

Capítulo 1

Preliminares.

1.1. Conceptos y Teoremas Básicos.

Los conceptos y teoremas que se observarán en este capítulo pueden encontrarse en [2],[13] y [17]. Se presentarán algunas definiciones, proposiciones y teoremas básicos del análisis convexo y la programación convexa que se necesitarán para una mejor comprensión de este trabajo. 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 $x = (x_1, \ldots, x_n)^t$, donde cada $x_i \in \mathbb{R}$, para $i = 1, \ldots, n$ y donde x^t denotará el vector traspuesto de x.

El producto interno de dos vectores x y y en $\mathbb{R}^n$ será representado por:

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

Si $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 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 $\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n : x > 0\}$ y $\mathbb{R}^n_{+} = \{x \in \mathbb{R}^n : x \geq 0\}$, donde $x > 0$ denotará que $x_i > 0$, para todo $i \in \{1, \ldots, n\}$ y $x \geq 0$ denotará que $x_i \geq 0$, para todo $i \in \{1, \ldots, n\}$.

***Definición* 1.1.1.** *Un subconjunto M de $\mathbb{R}^n$ es llamado afín si y sólo si*

$$(1 - \lambda)x + \lambda y \in M,$$

para cada $x, y \in M, \lambda \in \mathbb{R}$.

Es decir, un conjunto es afín, cuando la recta que une a cualquier par de puntos del conjunto está totalmente contenido en éste. El conjunto vacío $\emptyset$ y $\mathbb{R}^n$ son ejemplos extremos de conjuntos afines.

***Definición* 1.1.2.** *Un subconjunto C de $\mathbb{R}^n$ es llamado convexo si y sólo si*

$$(1-\lambda)x + \lambda y \in C,$$

para cada $x, y \in C, \lambda \in (0,1)$.

Es decir, un conjunto es convexo, cuando el segmento de recta que une a cualquier par de puntos del conjunto está totalmente contenida en éste.

***Definición* 1.1.3.** *Sea C un conjunto convexo y no-vacío en $\mathbb{R}^n$. Una función $f : C \to \mathbb{R}$ 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* 1.1.4.** *Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 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 denotaremos a la clase de tales funciones por $Conv\mathbb{R}^n$.

***Definición* 1.1.5.** *El dominio efectivo de $f \in Conv\mathbb{R}^n$ es el conjunto no-vacío*

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

Las definiciones (1.1.3) y (1.1.4) pueden hacerse equivalentes extendiendo la función convexa f definida en (1.1.3), haciendo $f(x) = +\infty$ para x no perteneciente a C y así se tiene una nueva f, la cual está ahora en $Conv\mathbb{R}^n$. Recíprocamente, dada $f \in Conv\mathbb{R}^n$, puede hacerse $C = dom f$ y se obtiene así una función convexa en el sentido de la definición (1.1.3).

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

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

lo cual nos permite visualizar la siguiente definición.

***Definición* 1.1.6.** *Dada $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función no idénticamente igual a $+\infty$, el epígrafo de f es el conjunto no-vacío dado por:*

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

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

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

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

Las siguientes propiedades son equivalentes:

(i) $f \in Conv\mathbb{R}^n$.

(ii) $epi(f)$ es un conjunto convexo en $\mathbb{R}^n \times \mathbb{R}$.

(iii) "$epi(f)$ estricto" es un conjunto convexo en $\mathbb{R}^n \times \mathbb{R}$.

***Definición* 1.1.7.** *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función. Se llaman conjuntos de sub-nivel de f a los siguientes, para cada $r \in \mathbb{R}$;*

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

Obsérvese que:

- $(x,r) \in epi(f) \Leftrightarrow x \in S_r(f)$.
- Si $f \in Conv\mathbb{R}^n$, entonces $S_r(f)$ es convexo para cada $r \in \mathbb{R}$ (posiblemente vacío).
- Una función cuyos conjuntos de sub-nivel son todos convexos no necesariamente es convexa (ver Fig 1.1.1).

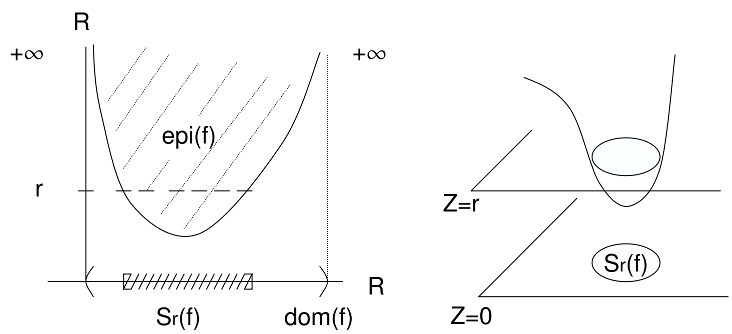

Fig.1.1.1 Función convexa, epígrafo, dominio efectivo y conjunto de sub-nivel.

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

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

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

***Definición* 1.1.8.** *Una función convexa $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ 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 $C = domf$ es no-vacío y la restricción de f a C es finita.*

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

***Teorema* 1.1.2.** *(ver[17], Teorema 1.1)*

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

***Teorema* 1.1.3.** *(ver[17], Teorema 1.5)*

Las transformaciones afines de $\mathbb{R}^n$ en $\mathbb{R}^m$ son funciones T de la forma $Tx = Ax + a$, donde A 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 $domf$ se denotará por $aff(domf)$ y se define como la intersección de todos los conjuntos afines que contienen a $domf$ [13].

- El interior relativo del epígrafo de f se denotará por:
  $ri(epi(f)) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : x \in ri(domf), r > f(x)\}$.

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

***Proposición* 1.1.2.** *(ver [13], Cap. IV, Prop. 1.2.1)*

Sea $f \in Conv\mathbb{R}^n$, entonces f es minimizada por alguna función afín; es decir, para todo $x_0 \in ri(domf) : \exists s \in L_{aff(domf)} / f(x) \geq f(x_0) + \langle s, x - x_0\rangle$ para todo $x \in \mathbb{R}^n$, donde $L_{aff(domf)}$ es el subespacio paralelo a $aff(domf)$.

Observe que la función afín puede ser forzada a coincidir con f en x_0.

***Definición* 1.1.9.** *Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es semicontinua inferior si para cada $x \in \mathbb{R}^n$ se cumple que:*

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

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

***Proposición* 1.1.3.** *(ver[13], Cap. IV, Prop.1.2.2)*

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

(i) f es semicontínua 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* 1.1.10.** *La función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 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* 1.1.11.** *La clausura de una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es la función $clf : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ definida por:*

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

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

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

***Proposición* 1.1.4.** *(ver[13], Cap. IV, Prop.1.2.6)*

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

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

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

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

Veamos algunas proposiciones sobre operaciones funcionales que preservan la convexidad.

***Proposición* 1.1.5.** *(ver[13], Cap.IV, Prop. 2.1.1)*

Sean $f_1, \ldots, f_n \in Conv\mathbb{R}^n$[respectivamente en $C\overline{onv}\mathbb{R}^n$], sean $t_1, ...t_n$ 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 $C\overline{onv}\mathbb{R}^n$].

***Proposición* 1.1.6.** *(ver[13], Cap.IV, Prop. 2.1.2)*

Sea $\{f_j\}_{j\in J}$ una familia arbitraria de funciones convexas [resp. convexas cerradas]. Si existe x_0 tal que $\sup_j\{f_j(x_0)\} < +\infty$, entonces la función siguiente:

$$f := \sup\{f_j : j \in J\},$$

es convexa [resp. convexa cerrada].

***Definición* 1.1.12.** *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia no necesariamente convexa, definimos la función conjugada f^* mediante:*

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

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

Por las proposiciones anteriores claramente $f^* \in \overline{Conv}\mathbb{R}^n$,

***Proposición* 1.1.7.** *(ver[13], Cap.IV, Prop. 2.1.5)*

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

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

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

Ejemplo: Sea $f \in Conv\mathbb{R}^n$, tómese $x_0 \in domf$, $d \in \mathbb{R}^n$ y defina $A : \mathbb{R} \ni t \mapsto A(t) = x_0 + td$, claramente A es afín y $f \circ A$ es por tanto convexa según la proposición anterior.

***Proposición* 1.1.8.** *(ver[13], Cap.IV, Prop. 2.1.8)*

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

Observe que de las proposiciones (1.1.5) y (1.1.7) se tiene que; si f es una función convexa y $u > 0$, entonces la función siguiente:

$$f_u : \mathbb{R}^n \ni x \mapsto f_u(x) = uf\left(\frac{x}{u}\right)$$

es también convexa y con ésta se puede definir definir la "perspectiva"de f como la función:

$$\widetilde{f} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$$

dada por

$$\widetilde{f}(u, x) := \begin{cases} uf(\frac{x}{u}) & si \quad u > 0 \\ +\infty & otro \quad caso. \end{cases}$$

***Proposición* 1.1.9.** *(ver[13], Cap.IV, Prop.2.2.1)*

Si $f \in Conv\mathbb{R}^n$ *, entonces su perspectiva* $\widetilde{f} \in Conv\mathbb{R}^{n+1}$.

***Definición* 1.1.13.** *Sea C un conjunto convexo, cerrado y no-vacío. Para cada $x \in C$, se define el cono asintótico en x como sigue:*

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

$C_\infty(x)$ 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* 1.1.10.** *(ver[13], Cap. III, Prop. 2.2.1)*

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

***Definición* 1.1.14.** *El cono asintótico o también llamado cono de recesión de un conjunto convexo cerrado C es el cono asintótico C_∞.*

Algunos conos asintóticos se observan en la figura 1.1.2

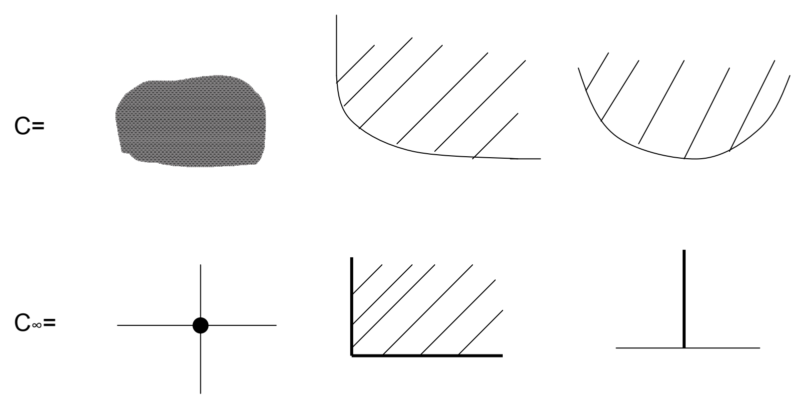

Fig.1.1.2 Algunos conos de recesión o conos asintóticos.

***Proposición* 1.1.11.** *(ver [13], Cap. III, Prop. 2.2.3).*

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

***Proposición* 1.1.12.** *(ver[13], Cap. IV, Prop. 3.2.2)*

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

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

donde x_0 es arbritrario en el dominio de f.

***Definición* 1.1.15.** *A esta función f'_∞ se le llama función de recesión.*

***Proposición* 1.1.13.** *(ver[13], Cap. IV, Prop.3.2.5)*

Sea $f \in C\overline{onv}\mathbb{R}^n$. Todos los conjuntos de sub-nivel de f, no-vacíos, tienen 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* 1.1.16.** *Las funciones $f \in C\overline{onv}\mathbb{R}^n$ que satisfacen (i), (ii), (iii) en la proposición anterior se llaman 0-coercivas. Equivalentemente las funciones 0-coercivas son aquellas que crecen al infinito, es decir;*

$$f(x) \to +\infty \quad \textit{cuando} \quad ||x|| \to +\infty,$$

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

Un importante caso particular es cuando se tiene que:

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

En este caso f crece al infinito más rápido que cualquier función afín y tales funciones son llamadas 1-coercivas. La siguiente figura muestra algunos ejemplos. 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 hay coercividad por un solo lado.

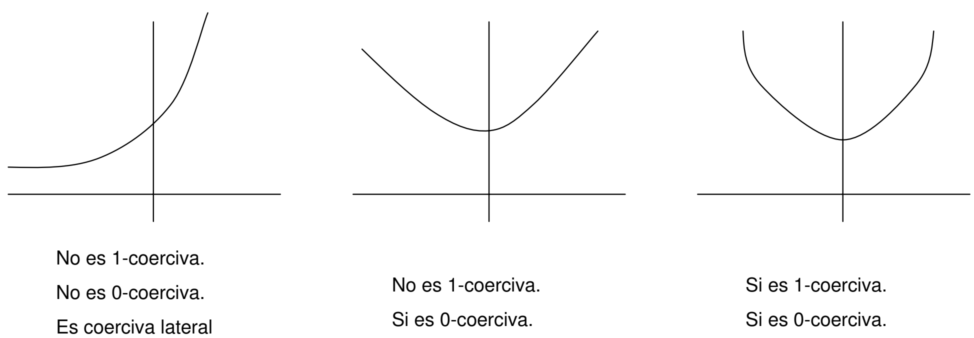

Fig.1.1.3 Coercividad.

***Definición* 1.1.17.** *Una dirección $d \neq 0$ es una dirección de recesión de f si y sólo si $f'_\infty(d) \leq 0$.*

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

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:

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

(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[13], Cap. IV, Teorema 4.3.1)*

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

***Definición* 1.1.19.** *Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa y propia es esencialmente suave, si satisface las siguientes condiciones:*

(i) $int(domf) \neq \emptyset$.

(ii) f es diferenciable en $int(domf)$.

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

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

***Definición* 1.1.20.** *Sea $f \in Conv\mathbb{R}$, 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[13], Cap. I, Teorema.4.1.1)*

Sea $f \in Conv\mathbb{R}$, 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* 1.1.21.** *Sean $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 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.*

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

se define la derivada direccional de f en x en la dirección d como:

$$f'(x, d) := \lim_{t \downarrow 0} q(t) = \inf\{q(t) : t > 0\}.$$

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

Por otro lado:

$$f'(x, -d) = \lim_{t \downarrow 0} \frac{f(x - td) - f(x)}{t} = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{-t}$$

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

***Definición* 1.1.22.** *El subdiferencial $\partial f(x)$ de f en x es el conjunto convexo, compacto y no-vacío de $\mathbb{R}^n$definido como:*

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

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

***Definición* 1.1.23.** *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 [13], Cap. IV, Teorema 4.3.1)*

Las definiciones (1.1.22) y (1.1.23) son equivalentes. La definición (1.1.23) significa que los elementos de $\partial f(x_0)$ son los vectores directores de los hiperplanos que pasan por $(x_0, f(x_0)) \in \mathbb{R}^n \times \mathbb{R}$.

***Proposición* 1.1.14.** *(ver[13],Cap. VI, Prop. 1.3.1)*

(i) Un vector $s \in \mathbb{R}^n$ es un subgradiente de f en x si y sólo si $(s, -1) \in \mathbb{R}^n \times \mathbb{R}$ es normal al epi(f) en $(x, f(x))$. En otras palabras:

$$N_{epi(f)}(x, f(x)) = \{(\lambda s, -\lambda) : s \in \partial f(x), \lambda \geq 0\}.$$

(ii) El cono tangente al epi(f) en $(x, f(x))$ es el epígrafo de la función derivada direccional $d \mapsto f'(x, d)$

$$T_{epi(f)}(x, f(x)) = \{(d, r) : r \geq f'(x, d)\}.$$

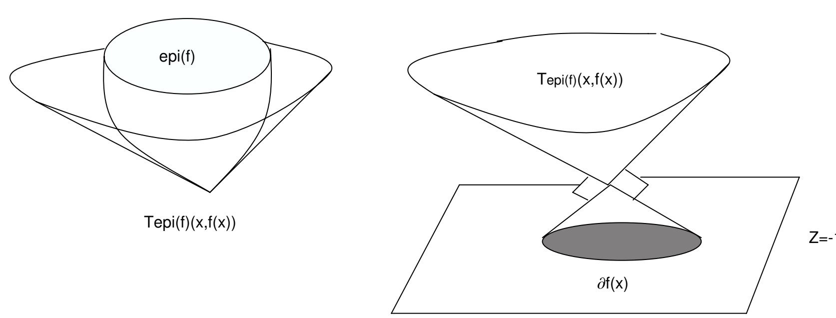

Fig.1.1.4 El conjunto subdiferencial.

En la figura 1.1.4 se observa lo presentado en el teorema anterior.

El siguiente teorema nos presenta condiciones de minimalidad equivalentes que se relacionan con la subdiferencial.

***Teorema* 1.1.8.** *(ver[13], Cap. VI, Teorema 2.2.1)*

Para $f : \mathbb{R}^n \to \mathbb{R}$ 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[13], Cap. VI, Teorema 4.4.1)*

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

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

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:

$$\min\{f(x) : x \in \mathbb{R}^n\}, \tag{1.2.1}$$

sujeto a,

$$h(x) = 0, \tag{1.2.2}$$

$$g(x) \leq 0. \tag{1.2.3}$$

donde $x = (x_1, \ldots, x_n)^t$ es el vector de las variables de decisión, $f : \mathbb{R}^n \to \mathbb{R}$ es la función objetivo, $h : \mathbb{R}^n \to \mathbb{R}^l$ y $g : \mathbb{R}^n \to \mathbb{R}^m$ 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 le llamará solución factible y el conjunto de todas las soluciones factibles se le llamará región factible.

El problema anterior en el caso en que todas las funciones son convexas se le 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 $f : \mathbb{R}^n \to \mathbb{R}$ 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 difrenciable) (1.2.1)-(1.2.2)-(1.2.3), si existe un par de vectores $\lambda \in \mathbb{R}^l$ y $\mu \in \mathbb{R}^m$ tales que:

$$\nabla f(\bar{x}) + \sum_{k=1}^{l} \lambda_k \nabla h_k(\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j(\bar{x}) = 0, \tag{1.2.4}$$

$$h_k(\bar{x}) = 0, \tag{1.2.5}$$

$$g_j(\bar{x}) \leq 0, \tag{1.2.6}$$

$$\mu_j g_j(\bar{x}) = 0, \tag{1.2.7}$$

$$\mu_j \geq 0. \tag{1.2.8}$$

A los vectores λ y μ se les llama multiplicadores de Lagrange o de Karush-Kuhn-Tucker. La condición (1.2.7) es conocida como condición de complementariedad, la condición (1.2.8) es conocida como condición de factibilidad dual y (1.2.5)-(1.2.6) son llamadas condiciones de factibilidad primal.

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

Debido a que la diferenciabilidad es un concepto local, esta permite caracterizar los mínimos locales, sin embargo, no es posible emplearla para caracterizar los mínimos globales de un programa no lineal. Con el fin de caracterizar los mínimos globales se considera la condición de convexidad y así 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[2], Cap. 3, Teorema 3.4.3)*

Sea $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ 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[2], Cap. 4, Teorema 4.3.7)*

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,* $(\bar{x}, \bar{\mu}, \bar{\lambda})$*, 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* $h_k(x)$*, 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

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

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

Se define la función Lagrangeana como

$$l(x, \mu, \lambda) = f(x) + \lambda^t h(x) + \mu^t g(x). \tag{1.2.9}$$

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

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

es decir, el problema dual viene dado por:

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

.

La teoría de dualidad es importante para la resolución de problemas no lineales, ver [2]. 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[2], Cap. 6, Teorema 6.2.1 y Corolarios 1,2,3 y 4)*

1. $sup\{\theta(\lambda,\mu) : \mu \geq 0\} \leq inf\{f(x) : h(x) = 0, g(x) \leq 0\}$ *(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* $\lambda \in \mathbb{R}^l$ *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 primal, lo cual permite resolver el problema primal mediante la resolución del problema dual. La idea básica es, por tanto, encontrar condiciones que garanticen la propiedad 1 del teorema anterior solamente con la condición de igualdad, como es el caso en los programas convexos. El siguiente teorema es un ejemplo que permite visualizar como actua esta propiedad de dualidad debil con solo condición de igualdad en el caso de la programación convexa.

***Teorema* 1.2.4.** *(ver[2], Cap. 6, Teorema 6.5.2)*

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* $l(x, \mu, \lambda) = f(x) + \lambda^t h(x) + \mu^t g(x)$ *sujeto a* $x \in X$.

Capítulo 2

Métodos de Punto Proximal y de Lagrangeano Aumentado.

En este capítulo se hace un breve resumen, siguiendo el estudio ofrecido en [9], de los métodos de punto proximal, en que consisten y como han evolucionado desde el concepto de regularización así como su relación con los métodos de Lagrangeano aumentado y los distintos tipos de familias de penalidades que generan distintas casi-distancias o medidas divergentes. Se destaca la evolución histórica de dichos métodos, los distintos cambios que han sido desarrollados en las distintas alternativas de penalización y finalmente los métodos de multiplicadores desarrollados en [6], los cuales sirvieron de inspiración para desarrollar una casi-distancia generalizada. La razón para revisar los distintos métodos de punto proximal, es que estos son utilizados para resolver el problema dual regularizando la función dual con diversas casi-distancias y al mismo tiempo servirá de marco teórico al estudio que se desarrollará en el capítulo 3. La intención en este trabajo es seguir creciendo en esa dirección, se presentará una familia de penalidades que generaliza a muchas de las conocidas actualmente y la cual, a través de shifts, genera una casi-distancia generalizada que incluye a las φ-divergencias y a las de núcleos homogéneos de segundo orden como casos particulares. Más aún, se generalizan éstos obteniéndose núcleos homogéneos no cuadráticos con potencias reales en los multiplicadores.

2.1. El concepto de regularización.

Dado un problema de la forma $L(f) = 0$ donde f es un elemento de algún conjunto $\mathbb{X}$ (usualmente un espacio de funciones) y $L : \mathbb{X} \to \mathbb{X}$ es un operador, se quiere reemplazar L por otro operador regularizado $L + \lambda M$ con $\lambda \in \mathbb{R}$ y $M : \mathbb{X} \to \mathbb{X}$ donde M sea tal, que el problema,

$$L(f) + \lambda M(f) = (L + \lambda M)(f) = 0,$$

tenga solución única f_λ para cada $\lambda > 0$ y si es posible, que f_λ se aproxime a una solución de $L(f) = 0$, cuando λ se aproxime a 0.

Ejemplo: Si $\mathbb{X} = \mathbb{R}^n$ y $L = \nabla f$ donde $f : \mathbb{R}^n \to \mathbb{R}$ es una función convexa, se considera el problema de encontrar x tal que:

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

o bien encontrar x tal que;

$$x \in arg\min\{f(x) : x \in \mathbb{R}^n\}. \tag{2.1.1}$$

Supóngase que $g : \mathbb{R}^n \to \mathbb{R}$ es 0-coerciva, convexa y f es acotada inferiormente. El problema (2.1.1) puede no tener solución o más de una solución, pero su regularización:

$$\min\{f(x) + \lambda g(x) : x \in \mathbb{R}^n\}, \tag{2.1.2}$$

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, así se tiene garantizada la existencia de soluciones. Por otro lado, la convexidad estricta, implica la unicidad de tal solución. Es decir, (2.1.2), bajo hipótesis adecuadas, tendrá una única solución $x(\lambda)$ tal que $\lim_{\lambda \to 0^+} x(\lambda)$ existe y resuelve (2.1.1).

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 $X = \mathbb{R}^n$ y $L = \nabla f$ donde $f : \mathbb{R}^n \to \mathbb{R}$ es una función convexa.

Diversos métodos se han desarrollado para sobrellevar la condición planteada anteriormente, al menos teóricamente, originándose así regularizaciones convenientes. Entre éstas se encuentran los métodos de punto proximal.

2.2. Método de Punto Proximal.

Este método resuelve el problema siguiente:

$\min\{f(x) : x \in \overline{X} \subset \mathbb{R}^n\}$, donde f es una función convexa y de clase C^1 en X, con $X \subset \mathbb{R}^n$ abierto y convexo y $\overline{X}$ es el conjunto clausura de X.

Considérese el algoritmo que genera una sucesión de la siguiente forma:

$$x^0 \in \mathbb{R}^n, \tag{2.2.1}$$

$$x^{k+1} = argmin_{x \in \mathbb{R}^n}\{f(x) + \lambda_k ||x - x^k||^2\}, \tag{2.2.2}$$

donde λ_k es un número real que satisface $0 < \lambda_k < \widetilde{\lambda}$ para algún $\widetilde{\lambda} > 0$ (se incluye el caso λ_k constante) y $\|.\|$ es la norma Euclideana.

Es posible demostrar que bajo hipótesis de convexidad y diferenciabilidad de clase C^1, la sucesión generada por (2.2.1) y (2.2.2) converge a un minimizador de f.

Concretamente el resultado que se obtiene es el siguiente:

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

Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función convexa y continuamente diferenciable. Suponga que el conjunto U de minimizadores de f sobre $\mathbb{R}^n$ es no-vacío. Entonces la sucesión $\{x^k\}$ generada por (2.2.1) y (2.2.2) converge a $x^* \in U$.

Este teorema es la versión clásica sobre la convergencia del método de punto proximal. Inicialmente en 1965, Moreau [25], ofrece esta regularización, concretamente propone la siguiente función:

$F(y) = argmin_{x\in\mathbb{R}^n}\{f(x) + \frac{1}{2}||x-y||^2\}$, que luego es modificada por Yosida, agregando el parámetro λ convirtiéndose en la regularización Moreau-Yosida conocida por la expresión siguiente:

$F_\lambda(y) = argmin_{x\in\mathbb{R}^n}\{f(x) + \frac{1}{2}\lambda||x-y||^2\}$.

Observe que el método de punto proximal clásico sustituye el parámetro λ por una sucesión $\{\lambda^k\}$ acotada y de números positivos. Martinet, [26] en 1970, introduce este método en el estudio de la programación convexa, y posteriormente Rockafellar[27], generaliza la técnica para operadores monótonos maximales.

Este método de punto proximal, debe ser visto primeramente como un método teórico, ya que en su aspecto práctico nos lleva a la resolución de una sucesión de subproblemas relativamente similares al original, que tendrían que ser resueltos con el apoyo de otros métodos; sin embargo, hay situaciones en las que es conveniente reemplazar un problema de minimización por una sucesión de subproblemas, como es el caso en que se tiene un problema original con restricciones y se convierte en una sucesión de subproblemas irrestrictos, situación que se presenta en el método de Lagrangeano aumentado, usando teoría de dualidad de Fenchel, como se verá más adelante.

***Definición* 2.2.1.** *Un operador $T : \mathbb{R}^n \to \mathbb{R}^n$ 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 $f : \mathbb{R}^n \to \mathbb{R}$ diferenciable y convexa. Es claro que el gradiente sería sustituido por subgradiente en el caso en que la función no fuese diferenciable.

***Definición* 2.2.2.** *Un operador $T : \mathbb{R}^n \to \mathbb{R}^n$ 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 $T = T'$.

En el caso del método de punto proximal clásico, para que $\{x^{k+1}\}$ resuelva (2.2.2), haciendo $f_k(x) = f(x) + \lambda_k \|x - x^k\|^2$, se cumple:

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

o bien,

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

lo que sugiere una extensión natural a operadores monótonos dada por:

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

lo cual es equivalente a:

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

El resultado que muestra la convergencia del método de punto proximal para operadores monótonos es el siguiente:

***Teorema* 2.2.2.** *(ver[9], Teorema 4.2)*

Sea $T : \mathbb{R}^n \to \mathbb{R}^n$ un operador monótono maximal, suponga que existe $\bar{x}$ tal que $0 \in T(\bar{x})$, entonces la sucesión $\{x^k\}$ generada por $x^0 \in \mathbb{R}^n$ y

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

converge a x^*, tal que $0 \in T(x^*)$, donde $\{\lambda_k\}$ es positiva y acotada superiormente,

En la decada de los 90, surgen diferentes ideas para sustituir los núcleos cuadráticos por otros no cuadráticos, Attouch y Wets [28] consideran núcleos de la forma $k(z - x)$ donde $k : \mathbb{R}^n \to \mathbb{R}_+$ es una función estrictamente convexa y coerciva tal que $k(0) = 0$, recuperando con esta técnica la regularización Moreau-Yosida. Tambien Humes-Da Silva [29] consideran núcleos similares, con k estrictamente convexa y diferenciable en el origen, mostrando que todo punto de acumulación de la sucesión generada por sus métodos es solución óptima del problema que resuelve el método de punto proximal.

2.3. Lagrangeano Aumentado.

Como se expresó anteriormente, el algoritmo de punto proximal debe ser visto más como una herramienta teórica que como un algoritmo implementable. En general $f_k(x) = f(x) + \lambda_k||x^k - x||^2$ es más fácil de minimizar que f pero en cada iteración se requiere la minimizaciòn de una función sobre $\mathbb{R}^n$ para la cual otro método numérico debe ser usado, de manera que las bondades de este método se observarán cuando los subproblemas sean sustancialmente más fáciles de resolver que el original, al reemplazar un problema de minimización restringida por una sucesión de problemas de minimización irrestricta. Esto se acostumbra en diferentes situaciones, como por ejemplo, en el caso del método de Lagrangeano aumentado donde es posible probar que este método de Lagrangeano aumentado no es más que un caso particular del método de punto proximal, visto desde un punto de vista dual, garantizando también la convergencia de este método.

El problema a considerar es:

$$f^* = inf\{f(x) : g_i(x) \leq 0 \quad i = 1, \ldots, m\}, \tag{2.3.1}$$

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

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

$$L(x, \mu) = \begin{cases} f(x) + \sum_{i=1}^{m} \mu_i g_i(x) & si \quad y \geq 0 \\ +\infty & otro\ caso \end{cases} . \tag{2.3.2}$$

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 $x^k = argminL(x, \mu^k)$,con $x \in \mathbb{R}^n$, actualizando μ^k de alguna manera conveniente. Computacionalmente los problemas están relacionados con la discontinuidad de $L(x, .)$, surgiendo así la necesidad de mejorarlos. Para ello, se desarrolló el método de Lagrangeano aumentado.

Con el método de Lagrangeano aumentado se obtienen mejores propiedades de convergencia.

Una descripción de tal método es la siguiente:

Para $x \in \mathbb{R}^n$, sea $x^+ \in \mathbb{R}^n$ definido como $x_j^+ = max\{x_j, 0\}$. Tómese $\alpha > 0$ y defínase el Lagrangeano aumentado clásico $L_\alpha : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ por:

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

$L_\alpha(x,\mu)$ es diferenciable si y sólo si f y g_i son diferenciables, aunque L_α no es necesariamente dos veces diferenciable cuando 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 de Lagrangeano aumentado clásico genera dos sucesiones: $\{x^k\} \subset \mathbb{R}^n$ y $\{\mu^k\} \subset \mathbb{R}^m$ dadas así:

Dados

$$\mu^0 \in \mathbb{R}^m_+, \tag{2.3.3}$$

$$x^k = argmin L_\alpha(x,\mu^k), x \in \mathbb{R}^n, \tag{2.3.4}$$

$$\mu_i^{k+1} = [\mu_i^k + 2\alpha g_i(x^k)]^+. \tag{2.3.5}$$

Considere ahora la función dual $d(\mu) = \inf\{L(x,\mu) : x \in \mathbb{R}^n\}$ con L como en (2.3.2). La función d es cóncava (ver [23]).

El siguiente teorema demuestra que la sucesión $\{\mu^k\}$ generada por (2.3.3) y (2.3.5) es la misma sucesión generada por el método de punto proximal aplicada a la función $-d$.

***Teorema* 2.3.1.** *(ver[9], Teorema 7.1)*

Sea $\{\overline{\mu}^k\} \subset \mathbb{R}^m$ la sucesión generada por el método de punto proximal para $\min\limits_{\mu\in\mathbb{R}^m}\{-d(\mu)\}$, con $\lambda_k = \frac{1}{4\alpha}$ y $\{\mu^k\}$ la sucesión dada por (2.3.3) y (2.3.5). Si $\overline{\mu}^0 = \mu^0$ entonces $\overline{\mu}^k = \mu^k$ para todo k.

Al combinar el Teorema 2.3.1 y (2.3.5) se tiene garantía de que la sucesión $\{\mu^k\}$ generada por el método de Lagrangeano aumentado converge a un maximizador μ^* de la función dual d, el cual es un multiplicador óptimo del tipo Karush-Khun Tucker para el problema (2.3.1).

La convergencia de la sucesión $\{x^k\}$ no puede ser obtenida inmediatamente de la teoría de punto proximal. De hecho, se requiere incluir hipótesis adicionales, como por ejemplo, además de otros, la condición de Slater; que exista $\overline{x}$ tal que $g_i(\overline{x}) < 0$ para todo $i = 1, \ldots, m$ 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 solución de (2.3.1).

2.4. Penalización.

Considere el problema:

$$f^* = inf\{f(x) : x \in \overline{S} \subset \mathbb{R}^n\}. \tag{2.4.1}$$

Para resolver (2.4.1) se puede considerar una función de penalidad g tal que $g(x) = \infty$, si x no pertenece a S. Para $r > 0$, el problema regularizado:

$$\min_{x \in \mathbb{R}^n} \{f(x) + rg(x)\} \tag{2.4.2}$$

tendrá su solución en $\overline{S}$ (si tiene solución). Si (2.4.1) tiene solución y g es escogida adecuadamente, entonces (2.4.2) tendrá una única solución $x(r)$ y bajo hipótesis adicionales es posible probar que $\lim_{r \to 0} x(r)$ existe y resuelve (2.4.1). Como en la regularización, aquí también puede ocurrir que el problema (2.4.2) esté mal condicionado para valores pequeños de r. Es decir cuando $x \in \partial\overline{S}$ y $g(x)$ es muy grande, $rg(x)$ puede ser que no se comporte adecuadamente cuando $r \to 0$. 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 $x(\overline{r})$, obtenida para (2.4.2), resuelva (2.4.1) para algún $\overline{r}$ fijo y aunque la existencia teórica de tal $\overline{r}$ sea posible, el valor de $\overline{r}$ no tiene por qué ser conocido, sin embargo el mal comportamiento comentado arriba pudiera permanecer.

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.

Una función de penalidad bien conocida es la función de penalidad exacta definida como:

$$y \in \mathbb{R} \mapsto y^+(y) = max\{0, y\},$$

sin embargo, para suavizar este tipo de penalidad, se construyen funciones de penalidad generalizadas del tipo siguiente:

$$y \in \mathbb{R} \mapsto h(y) = r\theta\left(\frac{y}{r}\right),$$

donde $r \in (0,1]$ y $\theta : \mathbb{R} \to \mathbb{R}$ es convexa, estrictamente creciente y de clase C^2 y como se vió en el capítulo 1, así h es convexa. Básicamente se tienen dos tipos de éstas penalidades generalizadas, las que podemos llamar del tipo I (0-coercivas), que además de cumplir las condiciones mencionadas anteriormente, también satisfacen que $dom\theta = \mathbb{R}$, $\lim_{y\to-\infty} \theta^{'}(y) = 0$ y $\theta'_\infty(1) = 1$ y las del tipo II (1-coercivas), las cuales cumplen que $dom\theta = (-\infty, b)$, para $0 < b \leq +\infty$, $\theta'_\infty(-1) = 0$ y $\theta'_\infty(1) = +\infty$. A las penalidades del tipo II, se le agregarán otras condiciones para construir la casi-distancia generalizada que se desarrollará en este trabajo.

Por otro lado, los métodos de Lagrangeano aumentado han sido mejorados utilizando funciones dos veces diferenciables que aprovechan el uso de métodos tipo Newton. Estos generan básicamente dos sucesiones $\{x^k\} \in \mathbb{R}^n$ y $\{\mu^k\} \in \mathbb{R}^m$ dadas por:

$$x^{k+1} \in argmin_{x\in\mathbb{R}^n}\{L_{r^k}(x, \mu_k)\}, \tag{2.4.3}$$

$$\mu^{k+1} = \nabla P_{r^k}(x^{k+1}, \mu_k), \tag{2.4.4}$$

donde $L_{r^k}(x,\mu) = f_0(x) + P_{r^k}(\mu, f(x))$ es la función Lagrangeana aumentada, $r^k \in (0,1]$ y P es una apropiada función de penalidad que en lugar de la penalidad y^+, utiliza ahora la penalidad θ. Diferentes funciones de Penalidad han sido utilizadas en aproximaciones de Lagrangeano aumentado y es bien conocido que las sucesiones generadas por (2.2.2) y (2.4.4) son las mismas cuando se comienza con el mismo vector μ^0 según apreciamos similarmente en el Teorema (2.3.1); así que la convergencia de (2.4.4) está garantizada debido a la convergencia de la sucesión asociada en el método de punto proximal. En este trabajo, se probará, que con penalidades generales y el uso del shift, se obtienen casi-distancias con núcleos homogéneos de orden $p \geq 0$ y que nuevamente las sucesiones resultantes en cada caso son las mismas al partir del mismo vector μ^0, quedando como casos particulares de este estudio, los núcleos de orden 1 y 2, desarrollados en [11] y [16].

Otras técnicas se desarrollan a través de las distancias de Bregman y las φ-divergencias, Censo-Zenios [12] en 1992, ofrece las D-funciones, mejorando las ya conocidas distancias de Bregman y posteriormente Iusem-Svaiter-Teboulle [11] en 1994, ofrecen las φ-divergencias. Todos ellos, sustituyen los núcleos cuadráticos del método de punto proximal por estas casi-distancias. Un breve resumen de los diferentes aspectos de estas casi-distancias se muestran a continuación.

***Definición* 2.4.1.** *Sea S un subconjunto no vacío de $\mathbb{R}^n$.*
Una función $D : S \times S \to \mathbb{R}_+$ con $S = int\overline{S}$ tal que: $D(x,y) = 0$ si y sólo si $x = y$ y $D(x,y) \geq 0$, para cada $x, y \in S$, es llamada una casi-distancia en S.

Se consideran varias clases de tales casi-distancias, entre ellas podemos destacar las asociadas a $\varphi-$divergencias y las Distancias de Bregman. En este trabajo se dará a conocer una casi-distancia generalizada, que incluye a las anteriores como casos particulares y se aportará una relación adicional entre ellas.

En ambos casos el método de punto proximal consiste en generar una sucesión $\{x^k\}$, con $x^0 \in S$ y $x^{k+1} = argmin_{x\in\mathbb{R}^n}\{f(x) + r_k D(x, x^k)\}$.

2.5. Distancias de Bregman y Método de Punto Proximal.

***Definición* 2.5.1.** *Sea $S \subset \mathbb{R}^n$ un conjunto abierto y convexo, y sea $\overline{S}$ su clausura. Se dirá que una función convexa $h : \overline{S} \to \mathbb{R}$ es una función de Bregman con distancia de Bregman $D_h : \overline{S} \times S \to \mathbb{R}$ dada por:*

$$D_h(x, y) = h(x) - h(y) - (\nabla h(y))^t(x - y), \tag{2.5.1}$$

si las siguientes condiciones se cumplen:

(B1) h es continuamente diferenciable sobre S.

(B2) h es estríctamente convexa y contínua sobre $\overline{S}$.

(B3) Para todo $\delta \in \mathbb{R}$, los conjuntos parciales de sub-nivel :

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

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

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

(B4) Si $\{y^k\} \subset S$ converge a y^*, entonces $D_h(y^*, y^k)$ converge a 0.

(B5) Si $\{x^k\} \subset \overline{S}$ y $\{y^k\} \subset S$ son sucesiones tales que: $\{x^k\}$ es acotada, $\lim_{k\to+\infty} y^k = y^*$ y $\lim_{k\to+\infty} D_h(x^k, y^k) = 0$, entonces $\lim_{k\to+\infty} x^k = y^*$.

S es llamada zona de h. Es fácil ver que: $D_h(x, y) \geq 0$, para todo $x \in \overline{S}$, para todo $y \in S$ y que además $D_h(x, y) = 0$ si y sólo si $x = y$.

Observe que (B4) y (B5) se cumplen automáticamente cuando $\{x^k\}$ y y^* están en S, como una consecuencia de (B1), (B2) y (B3), así que se necesitan chequear sólo éstas primeras, ver [9]. Esta definición tiene condiciones redundantes y puede ser redefinida de la siguiente forma, ver [14]:

***Definición* 2.5.2.** *Una función $h : \Lambda \subset \mathbb{R}^n \to \mathbb{R}$ es llamada función de Bregman sobre un conjunto no-vacío, abierto y convexo S, donde $\overline{S} \subseteq \Lambda$, si satisface las siguientes condiciones:*

(i) h es continua sobre $\overline{S}$.

(ii) h es estríctamente convexa sobre $\overline{S}$.

(iii) h es diferenciable sobre S.

(iv) Si $x \in \overline{S}$ y $\alpha > 0$, entonces $R^h_\alpha(x) = \{x \in S / D_h(x,z) \leq \alpha\}$ es acotado.

(v) Si $\{x^k\} \subset S$ es una sucesión convergente con límite $x^* \in S$, entonces el siguiente límite existe y

$$\lim_{k\to+\infty} \langle \nabla h(x^k), x^* - x^k \rangle = 0.$$

***Definición* 2.5.3.** *Se dice que una función h de Bregman es coerciva en la frontera si para $\{y^k\} \subset S$ tal que $\lim_{k\to+\infty} y^k = y \in \partial S$, entonces $\lim_{k\to+\infty} (\nabla h(y^k))^t (x - y^k) = -\infty$ para todo $x \in S$.*

***Proposición* 2.5.1.** *(ver [9], Prop. 9.1)*

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

(i) $D_h(x,y) - D_h(x,z) - D_h(z,y) = \langle \nabla h(y) - \nabla h(z), z - x \rangle$ para todo $x \in \overline{S}$ y para todo $y, z \in S$.

(ii) $\nabla_1 D_h(x,y) = \nabla h(x) - \nabla h(y)$ para todo $x, y \in S$.

(iii) $D_h(.,y)$ es estríctamente convexa para cada $y \in S$.

Las distancias de Bregman son usadas para generar un método de punto proximal.

El problema a resolver es:

$$f^* = \inf\{f(x) : x \in \overline{S} \subset \mathbb{R}^n\}, \tag{2.5.2}$$

con $S \subset \mathbb{R}^n$ abierto y convexo, $\overline{S}$ la clausura de S y f es convexa y continua sobre $\overline{S}$. El método de punto proximal con distancia de Bregman está definido por:

$$x^0 \in S, \tag{2.5.3}$$

$$x^{k+1} = argmin\{f(x) + \lambda_k D_h(x, x^k)\}, \tag{2.5.4}$$

donde $x \in \overline{S}$, h es una función de Bregman con zona S y λ_k satisface $0 < \lambda_k \leq \widetilde{\lambda}$ para algún $\widetilde{\lambda} > 0$.

El siguiente teorema presenta la convergencia del método.

***Teorema* 2.5.1.** *(ver[9], Teorema 10.1)*

Si el problema (2.5.2) tiene solución y h es coerciva en la frontera con respecto a S, entonces la sucesión $\{x^k\}$ generada por (2.5.3) y (2.5.4) converge a una solución x^* del problema (2.5.2).

Un ejemplo que debe resaltarse es el de la distancia entrópica o divergencia Kullback-Leibler. Se obtiene a partir de la función de Bregman $h : \mathbb{R}^n_{++} \to \mathbb{R}$ dada por $h(x) = \sum_{i=1}^n x_i log x_i$, extendiendo la continuidad al origen al definir $0log0 = 0$, así se obtiene la distancia de Bregman

$$D_h(x, y) = \sum_{i=1}^n (x_i log \frac{x_i}{y_i} + y_i - x_i).$$

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

2.6. $\varphi-$Divergencias y Método de Punto Proximal.

Aquí se estudia otra clase de casi-distancias, que son denotadas por $d_\varphi(\cdot,\cdot)$ y están definidas sobre el ortante positivo de $\mathbb{R}^n$. Sea $\varphi : \mathbb{R}^n_{++} \to \mathbb{R}$, convexa y de clase C^3 , con las siguientes condiciones:

$$\varphi(1) = \varphi'(1) = 0; \quad \varphi''(1) > 0; \quad \lim_{t\to 0^+} \varphi(t) = +\infty.$$

***Definición* 2.6.1.** *Si φ satisface las condiciones mencionadas anteriormente, entonces la aplicación $d_\varphi : \mathbb{R}^n_{++} \to \mathbb{R}$, definida por $d_\varphi(x,y) = \sum_{j=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right)$, se dice que es una $\varphi-$divergencia.*

Las siguientes propiedades se obtienen de la definición anterior:

***Proposición* 2.6.1.** *(ver [11], Proposición 2.1)*

(i) $d_\varphi(x,y) \geq 0$, para todo $x, y \in \mathbb{R}^n_{++}$.

(ii) $d_\varphi(x,y) = 0 \quad sii \quad x = y$.

(iii) Los conjuntos de nivel de $d_\varphi(.,y)$ son acotados para todo $y \in \mathbb{R}^n_{++}$.

(iv) Los conjuntos de nivel de $d_\varphi(x,.)$ son acotados para todo $x \in \mathbb{R}^n_{++}$.

(v) $d_\varphi(x,y)$ es conjuntamente convexa y estrictamente convexa en x.

(vi) $\lim_{k\to+\infty} d_\varphi(y,y^k) = 0$ sii $y^k = y$.

Algunos ejemplos clásicos de $\varphi-$divergencias, son los siguientes:

Ejemplo 2.6.1 Si $\varphi_1(t) = t\log t - t + 1$, entonces $d_{\varphi_1}(x,y) = \sum_{j=1}^{n}\left(x_j \log \frac{x_j}{y_j} + y_j - x_j\right)$. Observe que esta es la casi-distancia Kullback-Leibler, la cual fue obtenida anteriormente vía distancias de Bregman.

Ejemplo 2.6.2 Si $\varphi_2(t) = t - \log t - 1$, entonces $d_{\varphi_2}(x,y) = d_{\varphi_1}(y,x)$.

Ejemplo 2.6.3 Si $\varphi_3(t) = (\sqrt{t}-1)^2$, entonces $d_{\varphi_3}(x,y) = \sum_{j=1}^{n}(\sqrt{x_j} - \sqrt{y_j})^2$.

Con las $\varphi-$divergencias se ha desarrollado un método de punto proximal, el cual se presenta a continuación:

Se quiere resolver el siguiente problema:

$$f^* = inf\{f(x) : x \geq 0\}, \tag{2.6.1}$$

con $f : \mathbb{R}^n \to \mathbb{R}$ convexa.

El método genera una sucesión $\{x^k\} \subset \mathbb{R}^n$ dada por:

$$x^0 > 0,$$

$$x^{k+1} = argmin\{f(x) + \lambda_k d_\varphi(x, x^k)\}, \tag{2.6.2}$$

con λ_k satisfaciendo $0 < \lambda_k \leq \widetilde{\lambda}$, para algún $\widetilde{\lambda}$. La condición de optimalidad para esta sucesión viene dada por: $u^k \in \partial f(x^{k+1})$ con $u_j^k = -\lambda_k \varphi'\left(\frac{x_j^{k+1}}{x_j^k}\right)$.

Si f es diferenciable, x^{k+1} es la solución x del sistema $\nabla_j f(x) + \lambda_k \varphi'(\frac{x_j}{x_j^k}) = 0$, el cual es un sistema de n-ecuaciones en las n-variables desconocidas $x_1, .., x_n$. Como se mencionó anteriormente, si el problema (2.6.1) tiene solución, la minimización en (2.6.2) se reduce a subconjuntos compactos ($\lim_{k\to+\infty} \varphi(t) = \infty$) por las condiciones impuestas sobre φ y así se garantiza la existencia de x^{k+1}. La unicidad de x^{k+1} se sigue de la convexidad de f y la proposición (2.6.1)(v), lo que permite probar que está bien definida la sucesión, sin embargo, la prueba de convergencia es mucho más dura que en los casos anteriores debido a que la sucesión no es Fejer convergente (ver [9]) al conjunto de soluciones de (2.6.1) y por ello se relaja la condición Fejer convergente a casi-Fejer convergente (ver [9]).

***Definición* 2.6.2.** *Una sucesión* $\{y^k\} \subset \mathbb{R}^n_{++}$*, es casi-Fejer convergente al conjunto* $U \subset \mathbb{R}^n_{++}$*, con respecto a la* $\varphi-$*divergencia* d_φ*, si para cada* $u \in U$*; existe una sucesión* $\{\varepsilon_k\} \subset \mathbb{R}_{++}$ *tal que:* $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$ *y para todo* $k \geq 0$*:*

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

***Proposición* 2.6.2.** *(ver [9], Prop. 13.1)*

Si $\{y^k\} \subset \mathbb{R}^n_{++}$ *es casi-Fejer convergente a* $U \subset \mathbb{R}^n_{++}$*, con respecto a la* $\varphi-$*divergencia* d_φ*, entonces* $\{y^k\}$ *es acotada. Si un punto clausura* $\overline{y}$ *de* $\{y^k\}$ *pertenece a* U*, entonces* $\overline{y} = \lim_{k\to+\infty} y^k$.

Los principales resultados para el método de punto proximal y $\varphi-$divergencias pueden encontrarse en [11] y son los siguientes:

***Proposición* 2.6.3.** *(ver [11], Proposición 4.1)*

La sucesión generada por el algoritmo de punto proximal con $\varphi-$*divergencias está bien definida y contenida en* $\mathbb{R}^n_{++}$*. Más aún* $u^k \in \partial f(x^{k+1})$ *para todo* $k \geq 0$.

***Proposición* 2.6.4.** *(ver [11], Proposición 4.2)*

Para todo $k \geq 0$ *se tiene que:*

$0 \leq \lambda_k d_\varphi(x^{k+1}, x^k) \leq \lambda_k d_{\widehat{\varphi}}(x^{k+1}, x^k) \leq f(x^k) - f(x^{k+1})$ *donde* $\widehat{\varphi}(t) = (1-t)\varphi'(t)$.

Corolario 2.6.1 (ver [11], Corolario 4.1)

(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) $\lim_{k\to+\infty} x_j^k \varphi\left(\frac{x_j^{k+1}}{x_j^k}\right) = 0$, para todo $j = 1, ..., n$.

***Proposición* 2.6.5.** *(ver [11], Proposición 4.3)*

Si la sucesión $\{x^k\}$ converge a x^* y si se cumple: (i) f es $C^1(\mathbb{R}^n_+)$ o (ii) $\exists z \in T$ tal que $I(x^*) \subset I(z)$ entonces $x^* \in T$; donde $T = \{x/x$ es solución óptima $\}$ y $I(x) = \{j \in \{1,...,n\} : x_j = 0\}$.

***Teorema* 2.6.1.** *(ver [11], Teorema 4.1)*

Si la sucesión $\{y^k\}$ es casi-Fejer convergente a un conjunto no-vacío V , entonces $\{y^k\}$ es acotada y $\{d(v, y^k)\}$ es acotada para todo $v \in V$.

El resultado final en [11] que demuestra la convergencia del método es el siguiente:

***Teorema* 2.6.2.** *(ver [11], Teorema 4.2)*

Si se cumple cualquiera de las siguientes condiciones:

(i) La sucesión $\{x^k\}$ tiene puntos límite.

(ii) El conjunto T es acotado.

(iii) Existe solución z del problema primal (2.6.1) tal que $\{x^k_j\}$ es acotada, para aquellas 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 (2.6.1), entonces la sucesión $\{x^k\}$ generada por (2.6.2) converge.

Más aún, bajo cualquiera de las condiciones (iii), (iv) o (v) o si f es $C^1(\mathbb{R}^n_+)$, cualquier punto límite de la sucesión es una solución del problema (2.6.1).

Corolario 2.6.2 (ver [11], Corolario 4.4)

Si $\varphi'(t) \leq \varphi''(1) \log t, \forall t > 0$, entonces la sucesión $\{x^k\}$ generada por (2.6.2) converge a una solución del problema (2.6.1).

Observe que la distancia entrópica Kullback- Leibler puede ser obtenida tanto de las φ-divergencias como de 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. También es conveniente resaltar que la casi-distancia generalizada sugerida en este trabajo, no requiere de la condición $\varphi'(t) \leq \varphi''(1)\log t, \forall t > 0,$ condición necesaria entre las hipótesis para obtener la convergencia de la sucesión generada por el algoritmo propuesto por la casi-distancia en [11].

2.7. Convexidad y Dualidad

En esta sección, se considerará el problema no lineal primal (P) definido a continuación con hipótesis de convexidad. Se definirá el problema dual (D) asociado a (P) y se formularán los conceptos básicos para así introducir el teorema de dualidad de Fenchel, que se utiliza como enlace entre los métodos de Lagrangeano aumentado y los métodos de punto proximal. Estos resultados pueden ser encontrados en [17], [13], [20]. Se hará un resumen de diversas familias de penalidad y resultados principales considerados en los métodos de Lagrangeano aumentado.

Considere el problema:

$$\text{(P)} \quad \begin{array}{l} \min \quad f_o(x) \\ f_i(x) \leq 0 \quad i = 1, ..., m \\ \quad x \in \mathbb{R}^n, \end{array}$$

donde $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, para $i = 0, ..., m$ son funciones convexas, propias y cerradas. Vamos a suponer a lo largo de esta sección que:

(A1) El conjunto de soluciones óptimas S^* es no-vacío y compacto.

(A2) Existe $\overline{x} \in dom(f)$ *tal que* $f_i(\overline{x}) < 0$, para $i = 1, ..., m$ (Condición de Slater).

El problema convexo dual asociado a (P) está dado por:

$$(D)\quad \begin{array}{ll} \min & -d(\mu) \\ s.a. & \mu \geq 0 \end{array}$$

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

Observe que $d(\cdot)$ es cóncava y así $-d(\cdot)$ es convexa.

Denotemos por $\widehat{f}$ y $\widehat{d}$ los valores óptimos de los problemas (P) y (D) respectivamente.

Es conocido que bajo la condición de Slater, el conjunto de soluciones óptimas de (D) es no vacío y compacto (ver [13]) y $\widehat{f} = \widehat{d}$, además, para todo $\overline{d} < \widehat{d}$, el conjunto de sub-nivel:

$$\{\mu \in \mathbb{R}^m | \mu \geq 0 \ y \ d(\mu) \geq \overline{d}\} \tag{2.7.1}$$

es compacto.

***Definición* 2.7.1.** *Un vector μ^* es un multiplicador de Lagrange para el problema (P) si $\mu^* \geq 0$ y*

$$d(\mu^*) = \inf_{x \in \mathbb{R}^n} l(x, \mu^*) = \widehat{f}.$$

Recuerde que la conjugada de f, denotada por f^*, es cerrada y convexa y además f^* es propia si y sólo si f es propia. La conjugada g^* de una función cóncava cerrada $g : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ es definida como:

$$g^*(s) = \inf_{x \in \mathbb{R}^n} \{x^t s - g(x)\}.$$

La función g^* es cóncava y cerrada. Observe que g^* no está definida como $-f^*$ para $f = -g^*$, en efecto, definiendo $f = -g$ se obtiene:

$$f^*(s) = \sup_{x \in \mathbb{R}^n} \{x^t s - f(x)\} = \sup_{x \in \mathbb{R}^n} \{-(-x^t s - g(x))\} = -\inf_{x \in \mathbb{R}^n} \{-x^t s - g(x)\} = -g^*(-s).$$

La siguiente proposición será utilizada para calcular la conjugada de la función de penalidad que será definida en este trabajo con la que se obtendrá la casi-distancia generalizada.

***Proposición* 2.7.1.** *(ver[20], Prop. X 1.3.1)*

Considere las funciones convexas f, f_j con $1 \leq j \leq m$. Entonces se verifican las siguientes afirmaciones:

1. *Si $g(x) = f(x) + \alpha$, entonces $g^*(s) = f^*(s) - \alpha$.*
2. *Si $g(x) = \alpha f(x)$ con $\alpha > 0$, entonces $g^*(s) = \alpha f^*(\frac{s}{\alpha})$.*
3. *Si $g(x) = f(\alpha x)$ con $\alpha \neq 0$, entonces $g^*(s) = f^*(\frac{s}{\alpha})$.*
4. *Si A es un operador lineal invertible, entonces $(foA)^* = f^*o(A^{-1})^*$.*
5. *Si $g(x) = f(x - x_0)$, entonces $g^*(s) = f^*(s) + \langle s, x_0\rangle$.*
6. *Si $g(x) = f(x) + \langle s_0, x\rangle$, entonces $g^*(s) = f^*(s - s_0)$.*
7. *Si $f_1 \leq f_2$, entonces $f_1^* \geq f_2^*$.*
8. *Si $dom f_1 \cap dom f_2 \neq \emptyset$ y $\alpha \in (0, 1)$, entonces*
$$(\alpha f_1 + (1 - \alpha)f_2)^* \leq \alpha f_1^* + (1 - \alpha)f_2^*.$$
9. *Si $f(x) = \sum_{j=1}^{m} f_j(x_j)$, entonces $f^*(s_1, ..., s_m) = \sum_{j=1}^{m} f_j^*(s_j)$.*

***Proposición* 2.7.2.** *(ver[20], Corolario X 1.4.4)*

Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función convexa, propia y cerrada, las siguientes equivalencias son satisfechas para $x, s \in \mathbb{R}^n$:

$$x \in \partial f^*(s) \Leftrightarrow s \in \partial f(x) \Leftrightarrow f(x) + f^*(s) - s^t x = 0. \tag{2.7.2}$$

Teorema* 2.7.1. *Dualidad de Fenchel

Sea f una función convexa y propia sobre $\mathbb{R}^n$, y sea g una función cóncava y propia sobre $\mathbb{R}^n$, sean f^* la conjugada convexa de f y g^* la conjugada cóncava de g.

Se verifica que:

$$\inf_{x\in\mathbb{R}^n}\{f(x)-g(x)\} = \sup_{s\in\mathbb{R}^n}\{g^*(s)-f^*(s)\}, \tag{2.7.3}$$

si alguna de las siguientes condiciones es satisfecha:

(a) $ri(domf)\cap ri(domg)\neq\emptyset$.

(b) f *y* g *son cerradas y* $ri(domg^*)\cap ri(domf^*)\neq\emptyset$.

Si (a) es satisfecha, el supremo es obtenido en algún s, si (b) es satifecha, el ínfimo es obtenido en algún x; si (a) y (b) son satisfechas simultáneamente, el ínfimo y el supremo son necesariamente finitos.

En la actualidad, existe una fuerte tendencia a estudiar los métodos de Lagrangeano aumentado motivado por su gran desempeño en diversos problemas prácticos y sobre todo por la elegante conexión teórica con los métodos de punto proximal via dualidad de Fenchel. Una excelente muestra en este sentido se puede encontrar en [21] y en las referencias citadas en el mismo, principalmente asociado a problemas con restricciones de desigualdades. No obstante, se puede decir que el interés en el estudio de Lagrangeanos aumentados fue iniciado por [22] extendiendo el método de [23] formulado originalmente para casos con igualdades.

Aunque el interés de esta tesis es trabajar en una metodología que involucra una familia de penalidades para el problema convexo, hagamos una breve revisión de las distintas familias de penalidades que se han desarrollado para los distintos problemas no lineales, con el fin de destacar algunos detalles trascendentes de la familia que se considerará en el capítulo 3.

Considere el problema con restricciones de igualdades:

$$\text{(P1)}\quad \begin{array}{l} \min \quad f_o(x) \\ f_i(x)=0 \quad i=1,...,m \\ x\in\mathbb{R}^n, \end{array}$$

donde las funciones $f_i : \mathbb{R}^n \to \mathbb{R}$ para $i = 0, ..., m$, son continuamente diferenciables.

La idea general de Hestenes-Powel fue usar una penalidad cuadrática para el Lagrangeano de este problema actualizando los multiplicadores en cada iteración.

Asi, se considera el Lagrangeano aumentado como una función:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto L_c(x, \mu) = l(x, \mu) + \sum_{i=1}^{m} c f_i(x)^2 = f_0(x) + \sum_{i=1}^{m} \mu_i f_i(x) + c f_i(x)^2,$$

donde $c > 0$ es el parámetro de penalización en el Lagrangeano de $(P1)$. Se resuelve el problema $(P1)$ a través de problemas irrestrictos que a partir de μ^0 generan sucesiones $\{x^k\}, \{\mu^k\}$ y $\{c^k\}$ dadas por:

$$x^{k+1} \in argmin\{L_{c^k}(x, \mu^k), x \in \mathbb{R}^n\} \ y$$

$$\mu_i^{k+1} = \mu_i^k + 2c^k f_i(x^k) \quad para \ i = 1, \ldots, m.$$

La regla de actualización de multiplicadores obedece a:

$$\nabla L_{c^k}(x^{k+1}, \mu^k) = \nabla l(x^{k+1}, \mu^{k+1}) = 0.$$

2.7.1. Familias de Penalidades para el Lagrangeano Aumentado

A continuación se muestra una familia de funciones de penalidades que son utilizadas por el metodo de Lagrangeano aumentado para un problema con restricciones de igualdades.

Familia de Penalidades (P_E) **descritas por Bertsekas** (ver [4])

Sea $\theta : \mathbb{R} \to \mathbb{R}$ tal que:

(i) θ es continuamente diferenciable y estríctamente convexa sobre $\mathbb{R}$.

(ii) $\theta(0) = 0$.

(iii) $\theta'(0) = 0$.

(iv) $\lim_{y \to -\infty} \theta'(y) = -\infty$

(v) $\lim_{y \to +\infty} \theta'(y) = \infty$

Ejemplos:

(a) $\theta(y) = \frac{1}{2}y^2$.

(b) $\theta(y) = \cos h(y) - 1$.

(c) $\theta(y) = \frac{1}{p}|y|^p, p > 1$.

El Lagrangeano aumentado usado en esta familia de funciones está dado por:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto L_r(x, \mu) = f_0(x) + \sum_{i=1}^{m} \mu_i f_i(x) + r \sum_{i=1}^{m} \theta\left(\frac{f_i(x)}{r}\right),$$

donde $r \in (0, 1]$. El método de multiplicadores correpondiente para resolver el problema $(P1)$ genera sucesiones $\{x^k\}, \{\mu^k\}$ y $\{r^k\}$ a partir de μ^0 tal que:

$$x^{k+1} \in argmin\{L_{r^k}(x, \mu^k), x \in \mathbb{R}^n\} \ y$$

$$\mu_i^{k+1} = \mu_i^k + \theta\left(\frac{f_i(x^k)}{r^k}\right), \ i = 1, \ldots, m.$$

Nuevamente la regla de actualización de multiplicadores satisface:

$$\nabla L_{r^k}(x^{k+1}, \mu^k) = \nabla l(x^{k+1}, \mu^{k+1}) = 0. \tag{2.7.4}$$

Para el caso con desigualdades, el primer enfoque fue dado por [22] y [23]. Considerando el problema (P), el Lagrangeano clásico aumentado surge naturalmente al considerar holguras en las restricciones. Se muestra, ver [4], que la minimización con respecto a estas holguras pueden ser escritas en términos de la variable x, obteniendo una expresión para L_c en términos de $x, \mu \ y \ c$ dada por:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto L_c(x, \mu) = f_0(x) + \frac{1}{2c} \sum_{i=1}^{m} \{max\{0, \mu_i + cf_i(x)\}^2 - \mu_i^2\} \tag{2.7.5}$$

y nuevamente la regla de actualización de multiplicadores de un proceso iterativo se puede obtener para (2.7.5) satisfaciendo (2.7.4) donde la sucesión de parámetros positivos $\{c_k\}$ es acotada.

La principal desventaja de esta función de penalidad, introducida por Rockafellar, está en que no tiene segundas derivadas continuas aunque las restricciones sean de clase C^2.

Familia de Penalidades (P_I) **descrita por Bertsekas** (ver [4])

Para un problema con desigualdades (P), Bertsekas considera una familia de penalidades:

$$y \in \mathbb{R}, \mu \in \mathbb{R}_+ \mapsto p(y, \mu) \in \mathbb{R},$$

tal que:

(i) p es continua en $\mathbb{R} \times \mathbb{R}_+$, continuamente diferenciable en $\mathbb{R} \times \mathbb{R}_{++}$ y existe $\lim\limits_{\mu \to 0^+} \dfrac{p(y, \mu) - p(y, 0)}{\mu}$, para cada $y \in \mathbb{R}$ y $p(\cdot, 0)$ es de clase C^1 con respecto a $y \in \mathbb{R}$.

(ii) $p(y, \cdot)$ es cóncava en $[0, +\infty)$, para todo y fijo.

(iii) Para todo $\mu > 0, p(\cdot, \mu)$ es convexa en $\mathbb{R}$ y satisface la condición:
Si $y_0 > 0$ y $\nabla_2 p(y_0, \mu) > 0$, entonces $p(y, \mu) - p(y_0, \mu) > (y - y_0)\nabla_2 p(y_0, \mu)$, para todo $y \neq y_0$. (convexidad estricta)

(iv) $p(0, \mu) = 0, \ \forall \mu > 0$.

(v) $\nabla_2 p(0, \mu) = \mu, \ \forall \mu > 0$.

(vi) $\lim\limits_{y \to -\infty} \nabla_y p(y, \mu) = 0, \forall \mu \geq 0$.

(vii) $\lim\limits_{y \to +\infty} \nabla_2 p(y, \mu) = +\infty, \forall \mu \geq 0$.

(viii) $\inf_{y \in \mathbb{R}} p(y, \mu) > -\infty, \forall \mu \geq 0$.

Dentro de esta familia de penalidades (P_I) está la subfamilia (P_E^+), caracterizada por funciones de la forma:

$$y \in \mathbb{R}, \mu \in \mathbb{R} \mapsto p(y,\mu) = \begin{cases} \mu y + \theta(y) & si \quad \mu + \theta'(y) \geq 0 \\ \min_{r \in \mathbb{R}}\{\mu r + \theta(y)\} & otro\ caso \end{cases}$$

donde θ pertenece a la familia (P_E) considerada para el caso de desigualdades. Para $\theta(y) = \frac{1}{2}y^2$ se recupera la función clásica de Rockafellar con su correspondiente Lagrangeano aumentado mostrado en (2.7.5), el cual se puede escribir para todo i como:

$$p(y_i,\mu_i) = \begin{cases} \mu_i y_i + \frac{1}{2}y_i^2 & si \quad \mu_i + y_i \geq 0 \\ -\frac{1}{2}\mu_i^2 & si \quad \mu_i + y_i < 0. \end{cases}$$

También se considera la familia de penalidades $(\widehat{P}_1)$ que utilizan los Lagrangeanos aumentados dos veces continuamente diferenciables que se decriben a continuación.

Familia de Penalidades $(\widehat{P}_I)$ **descrita por Bertsekas** (ver [4])

Sea $p : \mathbb{R}^2 \to \mathbb{R}$ de la forma $\mu\theta(y)$, donde $\theta : \mathbb{R} \to \mathbb{R}$ es de clase $C^2, \theta'(y) > 0$ para todo $y \in \mathbb{R}$ y:

(i) $\theta(0) = 0$.

(ii) $\theta'(0) = 1$.

(iii) $\lim_{y \to -\infty} \theta(y) > -\infty$.

(iv) $\lim_{y \to -\infty} \theta'(y) = 0$.

(v) $\lim_{y \to \infty} \theta'(y) = \infty$.

Considerando el problema (P) el Lagrangeano correspondiente está dado por:

$$\begin{aligned} x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto L_r(x,\mu) &= f_0(x) + r\sum_{i=1}^m p\left(\frac{f_i(x)}{r}, \mu_i\right) \\ &= f_0(x) + r\sum_{i=1}^m \mu_i\theta\left(\frac{f_i(x)}{r}\right). \end{aligned}$$

donde $p(y,\mu) = \mu\theta(y)$ y $r \in [0,1)$.

El algoritmo consiste en que a partir de un vector de multiplicadores inicial $\mu^0 > 0$, se generan sucesiones $\{x^k\}, \{\mu^k\}$ y $\{r^k\}$ tales que:

$$x^{k+1} \in argmin_x L_r^k(\cdot,\mu_k) \tag{2.7.6}$$

$$\mu_i^{k+1} = \mu_i^k \theta' \left(\frac{f_i(x^{k+1})}{r^k} \right) \ para \ i = 1, \ldots, m. \tag{2.7.7}$$

La sucesión $\{r^k\}$ es decreciente y positiva pudiendo ser constante a partir de un valor apropiado. Observe que $\mu^k > 0$, para todo k.

Se puede observar que la regla de actualización de multiplicadores satisface:

$$\nabla L_r(x^{k+1}, \mu^k) = \nabla l(x^{k+1}, \mu^{k+1}) = 0. \tag{2.7.8}$$

Como ejemplo de una penalidad de la familia $(\widehat{P}_I)$ se tiene el método de multiplicador exponencial determinado por:

$$P(y, \mu) = \mu(exp(x) - 1).$$

Tseng-Bertsekas [24] muestran la convergencia de la sucesión $\{\mu^k\}$ a una solución óptima del problema dual (D) y respecto de la sucesión primal $\{x^k\}$, se consigue probar convergencia en un sentido ergódico, esto es, la convergencia de la sucesión media dada por $y^k = \frac{c^k x^k + \cdots + c^0 x^0}{c^k + \cdots + c^0}$ y $c^k = \frac{1}{r^k}$ para todo k, así se abren las puertas para posteriores estudios relacionando los métodos de Lagrangeano aumentado con las casi-distancias.

Posteriormente, Polyak-Teboulle [16] consideran la familia de penalidades con características similares a la familia $(\widehat{P}_I)$ consiguiendo incluir otras penalidades conocidas.

Familia de Penalidades descrita por Polyak-Teboulle (ver [16])

Se considera el problema (P) con desigualdades del tipo $\geq$ y la familia de funciones θ tales que:

$\theta : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ estríctamente creciente, de clase C^2 con $dom\theta = (a, +\infty)$ y $a \in [-\infty, 0), \theta(a) = -\infty, \theta'(a) = +\infty$ satisfaciendo:

(i) $\theta(0) = 0$.

(ii) $\theta'(0) = 1$.

(iii) $\theta''(0) < 0$.

(iv) $\lim\limits_{y \to +\infty} \theta'(y) = 0$.

(v) $\lim\limits_{r \to 0^+} r\theta\left(\dfrac{y}{r}\right) = 0, \forall y > 0$.

Entre las penalidades incluidas en esta familia se encuentran:

(a) $\theta(y) = 1 - e^{-y}, y \in \mathbb{R}$ (método exponencial).

(b) $\theta(y) = \log(1 + y), y > -1$ (barrera modificada logarítmica).

(c) $\theta(y) = \frac{y}{1+y}, y > -1$ (barrera modificada hiperbólica).

(d) $\theta(y) = \begin{cases} \log(1+y) & si \quad y \geq \frac{1}{2} \\ -e^{-2y-1} + 1 - \log(2) & si \quad y \leq -\frac{1}{2} \end{cases}$

(barrera modificada con penalidad exponencial).

Dada θ satisfaciendo las propiedades (i)-(v) y $r \in (0, 1]$ observe que: $y \geq 0$ si y sólo si $r\theta(\frac{y}{r}) \geq 0$, luego las restricciones del problema (P) considerado, pueden ser expresadas equivalentemente como:

$$r\theta\left(\frac{f_i(x)}{r}\right) \geq 0, \quad i = 1, \ldots, m. \tag{2.7.9}$$

Considerando el Lagrangeano clásico correspondiente al minimizar f_0 sujeto a (2.7.9) se obtiene el Lagrangeano aumentado dado por:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto L_r(x, \mu) = f_0(x) + r \sum_{i=1}^{m} \mu_i \theta\left(\frac{f_i(x)}{r}\right).$$

El algoritmo usado sigue el mismo formato de (2.7.6) y (2.7.7). Esta técnica con respecto al Lagrangeano aumentado fue llamada principio de reescalamiento no lineal y conduce naturalmente a un problema dual que envuelve φ-divergencias.

Dada θ, se define la función φ como:

$$y \in \mathbb{R}_{++} \mapsto \varphi(y) = -\theta^*(y).$$

Un hecho importante es que φ satisface las condiciones necesarias que definen una φ-divergencia.

Una hipótesis hecha sobre las funciones θ consideradas es que θ' sea logarítmicamente convexa, o equivalentemente que θ' satisfaga:

$$\log\theta'(t) \geq \log\theta'(t_0) + \frac{\theta''(t_0)}{\theta'(t_0)}(t-t_0), \forall t > -1, t_0 \neq -1. \tag{2.7.10}$$

La condición anterior es usada para deducir la condición:

$$\varphi'(t) \leq \varphi''(1)\log(t), \forall t > 0,$$

usada en [11] para las pruebas de convergencia.

La casi-distancia asociada a esta familia de penalidades viene dada al definir $d_\varphi : \mathbb{R}^n_{++}\times\mathbb{R}^n_{++} \to \mathbb{R}$, como: $d_\varphi(x,y) = \displaystyle\sum_{j=1}^{n} y_j\varphi\left(\frac{x_j}{y_j}\right)$. Estas medidas divergentes son, como vimos anteriormente, conocidas como φ-divergencias. El principal resultado se refiere a la convergencia en un sentido ergódico de la sucesión $\{x^k\}$ generada por el Lagrangeano aumentado. Es decir que todo punto límite de la sucesión de medias $\overline{x}^s = \displaystyle\sum_{k=1}^{s}\frac{x^k}{s}$ es una solución óptima de (P).

Familia de Penalidades descrita por Ben-Tal y Zibulevsky (ver [3])

Considere el problema (P) y las funciones dadas por: $\theta : (-\infty, b) \to \mathbb{R}$ estrictamente convexas, estrictamente crecientes y de clase C^2 con $0 < b \leq \infty$ tales que:

(i) $\theta(0) = 0$.

(ii) $\theta'(0) = 1$.

(iii) $\lim_{y\to b}\theta'(y) = \infty$.

(iv) $\lim_{y\to -\infty}\theta'(y) = 0$.

(v) $\theta''(y) \geq \frac{1}{M}, \forall y \in [0,b]$, para algún $M > 0$.

En [3] se considera el Lagrangeano aumentado de la forma:

$$x \in \mathbb{R}^n, y \in \mathbb{R}^m \mapsto L_r(x,\mu) = f_0(x) + \sum_{i=1}^{m} \mu_i r_i \theta\left(\frac{f_i(x)}{r}\right), \tag{2.7.11}$$

donde $r_i = \pi(\mu_i)$ para $i = 1,\dots,m$ y $\pi : \mathbb{R}_{++} \to \mathbb{R}_{++}$ es una función decreciente y sublineal, es decir, $\forall \mu > 0 : \pi(\mu) \leq c\mu$ para algún $c > 0$.

El algoritmo genera sucesiones como en (2.7.6) y (2.7.7) usando la actualización:

$$r^{k+1} = \pi^k(\mu^{k+1}), \tag{2.7.12}$$

observe que en este caso r está en función de los multiplicadores, lo que fue sugerido en [24] para el método de multiplicadores exponencial. Con respecto a la sucesión primal, la misma tiene puntos límite que pertenecen al conjunto de soluciones óptimas de (P).

Entre las penalidades que satisfacen estos enfoques están:

(a) $\theta(y) = e^y - 1$. (método exponencial).

(b) $\theta(y) = -\log(1-y), -\infty < y < 1$. (barrera modificada).

(c) $\theta(y) = \begin{cases} y + \frac{1}{2}y^2 & si \quad y \geq -\frac{1}{2} \\ -\frac{1}{4}\frac{1}{1-y} - \frac{7}{6} & si \quad y \leq -\frac{1}{3} \end{cases}$ (cuadrática logarítmica).

(d) $\theta(y) = \begin{cases} y + \frac{1}{2}y^2 & si \quad y \geq -\frac{1}{3} \\ \frac{32}{27}\log(-2y) - \frac{3}{8} & si \quad y \leq -\frac{1}{3}. \end{cases}$ (recíproca cuadrática).

Considere nuevamente los problemas primal (P) y dual (D) con las hipótesis:

(A1) El conjunto de soluciones S^* de (P) es no vacío y compacto.

(A2) Existe $\overline{x} \in dom f$ tal que $f_i(\overline{x}) < 0, i = 1,\dots,m$. (Condición de Slater).

Sea:

$$\overline{f} = inf\{f_0(x) | f_i(x) \leq 0, i = 1,\dots,m\},$$

se considera la familia de funciones φ-divergencias que determinan la casi-distancia definida por:

$$d_\varphi(s,\mu) = \sum_{i=1}^{m} \mu_i \pi(\mu_i) \varphi\left(\frac{s_i}{\mu_i},\right), x \geq 0, y > 0 \tag{2.7.13}$$

(en nuestro contexto $\varphi = \theta^*$) donde la función:

$$\mu \in \mathbb{R}_+ \mapsto \pi(\mu) \in \mathbb{R}_+$$

es creciente y sublineal, es decir, $\forall \mu > 0$, se tiene que

$$\pi(\mu) \leq c\mu, \tag{2.7.14}$$

para algún $c > 0$. Considere la sucesión $\{\mu^k\}$ generada en el siguiente proceso:

$$\mu^{k+1} = argmax_x\{d(\mu) - d_\varphi(\mu, \mu^k)\}, u^0 > 0, \tag{2.7.15}$$

donde d es una función dual y d_φ fue definida en (2.7.13).

Se muestra, que esta sucesión es decreciente. Una condición adicional pedida es:

(A3) Sea d un vector en el conjunto de subgradientes de d_φ con respecto al primer argumento, es decir, $d \in \partial_s d_\varphi(s, \mu)$. Para todo $\epsilon > 0, \exists \delta > 0$ tal que si para algún $i \in \{1, \ldots, n\}, d_i > \epsilon$, entonces $d_\varphi(s, \mu) > \delta$.

Las condiciónes de sublinealidad de π en el siguiente lema son importantes para que (2.7.18) satisfaga la propiedad (A3).

***Lema* 2.7.1.1.** *(ver lema 2 en [3])*

Si $0 < y < x$ *y* $0 < \frac{\pi(y)}{y} < c$ *para algún* $c > 0$. *Entonces*

$$\rho(x, y) \geq \frac{1}{2}\frac{[\rho_1'(x, y)]^2}{cM}, donde\ 0 < M < +\infty.$$

El lema anterior inspira un lema que se utilizará para demostrar la convergencia del método que se propone en este trabajo con una casi-distancia generalizada, la cual se obtiene como la conjugada de una función de penalidad con shifts, y a pesar de que el lema propuesto en el capítulo 3 es similar en su enunciado a éste, su interpretación es muy diferente, además de que, la constante c involucrada en ambos tienen significados distintos, en este, c es cota de la función sublineal, en el capítulo 3 la constante c está definiendo la traslación de la función de penalidad propuesta.

El resultado obtenido bajo estas condiciones es el siguiente y es encontrado en [3].

***Proposición* 2.7.1.1.** *Supóngase que son satisfechas (A1) y (A2) y θ satisface las condiciones (i)-(v). Para cada k, considere π^k verificando la condición (2.7.14). Entonces las sucesiones $\{x^k\}$ y $\{\mu^k\}$ generadas por el método de Lagrangeano aumentado satisfacen:*

(i) $max\{0, f_i(x^k)\} \to 0,$ para $i = 1, \ldots, m.$

(ii) $f(x^k) \to \widehat{f}.$

(iii) $\mu_i^k f_i(x^k) \to 0,$ para $i = 1, \ldots, m.$

(iv) Las sucesiones $\{x^k\}$ y $\{\mu^k\}$ son acotadas y todos sus puntos límite, son soluciones de los problemas (P) y (D) respectivamente.

La condición (A3) es importante para la convergencia del método. La hipótesis (A2) garantiza que la sucesión de valores duales implícita en (2.7.15), es no decreciente; es decir, junto a (A3) tornan válida la afirmación (i) del teorema anterior, que junto al hecho de que la función π es sublineal conducen a (iii) y (ii).

Finalmente, observe que las familias de penalidades (P_I), presentadas por Polyak-Teboulle son consideradas por Ben-Tal-Zibulevsky y tienen en común la condición de que la función de penalidades debe crecer más rápido que una aproximación lineal. Podríamos llamar esta propiedad 1-coercividad lateral.

2.7.2. Relación entre los Métodos de Lagrangeano Aumentado y los Métodos de Punto Proximal

Los métodos de Lagrangeano aumentado y los de punto proximal están relacionados según se mostrará a continuación, Rockafellar [22] es el primero que observa dicha relación la cual puede apreciarse así:

Por una parte se tiene que:

$$
\begin{aligned}
d(\mu) &= inf\{f_0(x) + \mu^t f(x) : x \in \mathbb{R}^n\},\\
&= inf_{u\in\mathbb{R}^m} inf_{x\in\mathbb{R}^n}\{f_0(x) + \mu^t u, f(x) \le u\},\\
&= inf_{u\in\mathbb{R}^m}\{inf_{x\in\mathbb{R}^n}\{f_0(x), f(x) \le u\} + \mu^t u\},\\
&= inf_{u\in\mathbb{R}^m}\{v(u) + \mu^t u\},\\
&= -sup_{u\in\mathbb{R}^m}\{-\mu^t u - v(u)\},
\end{aligned}
$$

de donde $d(\mu) = -v^*(-\mu)$, donde $v(u) = inf_{x\in\mathbb{R}^n}\{f_0(x), f(x) \le u\}$ es la llamada función perturbación.

Por tanto:

$$
\begin{aligned}
inf_{x\in\mathbb{R}^n}\{f_0(x) + P_c(f_i(x), \mu)\} &= inf_{u\in\mathbb{R}^m}\{v(u) + P_c(u, \mu)\},\\
&= sup_{s\ge 0}\{-v^*(-s) - P_c^*(s, \mu)\},\\
&= sup_{s\ge 0}\{d(s) - P_c^*(s, \mu)\},\\
&= inf_{s\ge 0}\{-d(s) + P_c^*(s, \mu)\},
\end{aligned}
$$

donde $P_c(f_i(x), \mu)$ representa el lagrangeano penalizado.

Considere nuevamente el problema de interés (P) junto con su respectivo problema dual (D) y la función de Lagrangeano definida como:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto l(x,\mu) = \begin{cases} f_0(x) + \displaystyle\sum_{i=1}^{m} \mu_i f_i(x) & si \quad \mu \geq 0 \\ -\infty \qquad otro\ caso. \end{cases}$$

Teboulle [19] considera lagrangeanos aumentados a partir de casi-distancias.

Define:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto L_r(x,s) = sup_{s\in\mathbb{R}^m_+}\{l(x,s) - rD(s,\mu)\},$$

donde $r > 0$ y $D(s,\mu)$ es una distancia de Bregman. Por ejemplo en el caso en que $D(s,\mu)$ sea una φ divergencia, como se define en (3.10), se tiene:

$$L_r(x,\mu) = sup_{s\in\mathbb{R}^m_+}\{l(x,s) - r\sum_{i=1}^{m}\mu_i\varphi(\frac{s_i}{\mu_i})\}$$

$$= f_0(x) + \sum_{i=1}^{m} sup_{s_i\in\mathbb{R}_+}\{s_i f_i(x) - r\mu_i\varphi(\frac{s_i}{\mu_i})\}$$

$$= f_0(x) + r\sum_{i=1}^{m}\mu_i\varphi^*\left(\frac{f_i(x)}{r}\right),$$ (en nuestro contexto $\varphi = \theta^*$) y sobre hipótesis razonables, el método de multiplicadores es definido como:

Dada una sucesión $\{r^k\}$ de números positivos, genera sucesiones $\{x^k\} \subset \mathbb{R}^n$ y $\{\mu^k\} \subset \mathbb{R}^m$ tales que:

$$x^{k+1} \in argmin\{L^k_r(x,\mu_k) : x \in \mathbb{R}^n\}, \tag{2.7.16}$$

$$\mu^{k+1} = argmax\{(\mu, f_i(x^{k+1})) - r^k d_\varphi(\mu,\mu^k) : \mu > 0\}. \tag{2.7.17}$$

Observe que si μ^{k+1} satisface (2.7.17), entonces:

$$f_i(x^{k+1}) - r^k\varphi'\left(\frac{\mu_i^{k+1}}{\mu_i^k}\right) = 0, para\ i = 1,\ldots,m$$

para obtener,

$$\mu_i^{k+1} = \mu_i^k(\varphi')^{-1}\left(\frac{f_i(x^{k+1})}{r^k}\right)$$

o bien,

$$\mu_i^{k+1} = \mu_i^k (\varphi *)' \left(\frac{f_i(x^{k+1})}{r^k} \right), para\ i = 1, \ldots, m. \tag{2.7.18}$$

Esta última igualdad corresponde a la actualización de multiplicadores del método de Lagrangeano aumentado usado por Bertsekas [4] para la penalidad exponencial haciendo $\varphi(y) = y \log y$.

Posteriormente, Polyak y Teboulle [16] muestran que la conjugada de φ corresponde a la familia de penalidades para los métodos de Lagrangeano aumentado considerados en la sección 2.7.1 con desigualdades del tipo $\geq$.

Nótese que la actualización (2.7.18) es equivalente a:

$$\mu^{k+1} = argmax_{s \in \mathbb{R}_+^m} \{d(\mu) - r^k d_\varphi(\mu, \mu^k)\} \tag{2.7.19}$$

donde d es la función dual.

De esta manera, el iterado μ^{k+1} en (2.7.18) correspondiente a la actualización de los multiplicadores en el método de Lagrangeano aumentado considerado, también es solución del lado derecho de (2.7.19), que corresponde a la iteración k del método de punto proximal usando φ-divergencias para minimizar la función $-d(.)$ en el ortante positivo.

Se puede explicar aún más la relación anterior, obsérvese que, para el caso convexo y con hipótesis razonables, si x^{k+1} minimiza a $L_r^k(x, \mu_k)$ entonces x^{k+1} minimiza $l(x^{k+1})$, así.

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

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

$$\leq f_0(x^{k+1}) + \sum_{i=1}^m \mu_i f_i(x^{k+1})$$

$$= f_0(x^{k+1}) + \sum_{i=1}^m \mu_i^{k+1} f_i(x^{k+1}) + \sum_{i=1}^m (\mu_i - \mu_i^{k+1}) f_i(x^{k+1})$$

$$= d(\mu^{k+1}) + \sum_{i=1}^{m}(\mu_i - \mu_i^{k+1}) f_i(x^{k+1}).$$

Así, por ser d cóncava, se tiene:

$$f(x^{k+1}) \in \partial d(\mu^{k+1}) \tag{2.7.20}$$

donde $f(.) = (f_i(.), \ldots, f_m(.))$ y $\partial d(\mu)$, es el conjunto de subgradientes en μ de la función cóncava d (ver 2.7.4).

La actualización de multiplicadores dada por (2.7.18) es:

$$\mu_i^{k+1} = \mu_i^k (\varphi^*)' \left(\frac{f_i(x^{k+1})}{r^k} \right), \; para \; i = 1, \ldots, m,$$

y así usando la relación $(\varphi')^{-1} = (\varphi *)'$ se tiene para $i = 1, \ldots, m$

$$f_i(x^{k+1}) = r_k \varphi' \left(\frac{\mu^{k+1}}{\mu^k} \right). \tag{2.7.21}$$

De (2.7.20) y (2.7.21) obsérvese que:

$$r_k \begin{pmatrix} \varphi'(\frac{\mu_i^{k+1}}{\mu_i^k}) \\ \vdots \\ \varphi'(\frac{\mu_m^{k+1}}{\mu_m^k}) \end{pmatrix} \in \partial d(\mu^{k+1})$$

o equivalentemente:

$$0 \in \partial d(\mu^{k+1}) - r^k \begin{pmatrix} \varphi'(\frac{\mu_i^{k+1}}{\mu_i^k}) \\ \vdots \\ \varphi'(\frac{\mu_m^{k+1}}{\mu_m^k}) \end{pmatrix}. \tag{2.7.22}$$

Iusem, Svaiter y Teboulle [11] muestran un enfoque similar con ejemplos de funciones que satisfacen las relaciones anteriores. Con distancias de Bregman, además de el Trabajo de Teboulle [19], Eckstein [7] considera problemas con restricciones de igualdades para definir métodos de Lagrangeano aumentado en forma similar a

Teboulle [19]. Un estudio completo sobre la interrelación entre estas dos metodologías se puede encontrar en Iusem [21].

Humes y Da Silva [29] también relacionan el método de lagrangeano aumentado generalizado en el que $P_r(y, \mu)$ viene dada por:

$$P_r(y, \mu) = (r\theta(\cdot - \mu^k) + I^m_{\mathbb{R}_+}(\cdot))^*(y)$$

con el método de punto proximal via dualidad de Fenchel, donde I es la función indicadora.

Luego Kiwiel [14] para el problema $inf\{f(x) : x \in X\}$ de minimización convexa, generaliza las funciones de Bregman y cubre más aplicaciones que los métodos anteriores. Estas extensiones son llamadas B-funciones las cuales permiten el caso no diferenciable e infinito en la frontera de sus dominios efectivos. Básicamente supone que:

(i) f es una función convexa propia y cerrada.

(ii) $X \neq \emptyset$, cerrado y convexo.

(iii) h es una B-función.

(iv) $dom(f/X) \cap dom(h) \neq \emptyset$ donde $f/X = f + I_X$.

(v) $\{c_k\}$ tal que $c_k > 0$, para todo $k \geq 1$ y $\sum_{k=1}^{+\infty} c_k = +\infty$.

(vi) $\{\varepsilon_k\}$tal que $\varepsilon_k \geq 0$,para todo $k \geq 1$ y $\lim_{l\to\infty} \sum_{k=1}^{l} c_k\varepsilon_k / \sum_{k=1}^{l} c_k = 0$.

Estas condiciones adaptadas al problema dual le permiten generar el siguiente método: Para cada $k \geq 1$

$$\mu^k \in dom(f/X) \cap dom(\partial h),$$

$$y^k \in \partial h(\mu^k),$$

$$D_h^k(\mu, \mu^k) = h(\mu) - h(\mu^k) - \langle y^k, \mu - \mu^k\rangle,$$

$$\mu^{k+1} = argmin\{f/X + D_h^k(\mu, \mu^k)/c_k\}$$

Bajo estas condiciones el algoritmo descrito por Kiwiel aplicado al problema convexo primal dado por $inf\{f_0(x) : g_i(x) \leq 0,\ x \in \mathbb{R}^n\}$, se puede resumir en la iteración $k \geq 1$, de la siguiente forma; dados

$$\mu^k \in dom(-d)$$

$$y^k \in \partial h_+(\mu^k)$$

donde $h_+ = h + I_{\mathbb{R}^m_+}$ es una B-funcion, encontrar

$$x^{k+1} \in argmin_{x \in dom(f_0)}\{f_0(x) + \frac{1}{c_k} h^+(y^k + c_k g(x)) - h^+(y^k)\},$$

donde $h^+(.) = sup_{\mu \geq 0}\{\mu^t(.) - h(\mu)\}$ es la conjugada monótona (ver [17]) con actualización dada por:

$$\mu^{k+1} = \nabla h^+(y^k + c_k g(x^{k+1}))$$

con

$$y^k + c_k g(x^{k+1}) \in \partial h_+(\mu^{k+1}).$$

Observe que el método sugerido por Kiwiel plantea un shift variable que debe ser actualizado en cada iteración. En el capítulo 3 se desarrollará una casi-distancia generalizada que representa el aporte esencial de este trabajo y que a diferencia del método desarrollado por Kiwiel será generada por medio de un shift constante.

Finalmente Kiwiel presenta varios resultados sobre la convergencia de ambas sucesiones, la primal y la dual, para ello define entre otras, la familia de funciones siguiente: $\Phi_S = \{\phi : \mathbb{R} \longrightarrow (-\infty, +\infty]$ convexa propia cerrada y esencialmente suave con $intdom(\phi) = dom(\phi)$ y $\mathbb{R}_> \subset int\nabla\phi \subset \mathbb{R}$ y estrictamente convexa$\}$

***Proposición* 2.7.2.1.** *(ver [20], Teorema 9.18)*

Si $\phi \in \Phi_S, \quad inf\phi > -\infty, \quad argmaxd \neq \emptyset, \sum_{j=1}^{k} s_j\varepsilon_j/s_k \to 0, \sum_{k=1}^{\infty} c_k\varepsilon_k < \infty,$

$inf_k c_k > 0$ entonces $\mu^k \to \mu^\infty \in argmax\ d, d(\mu^k) \longrightarrow d^\infty$ y se cumple que: $\overline{\lim}_{k\to\infty} f(x^k) \leq d^\infty$ y $\overline{\lim}_{k\to\infty} g_i(x^k) \leq 0$ para $i = 1 \ldots m$, además si $\{x^k\}$ tiene punto límite x^∞, entonces x^∞ resuelve el problema primal y $f(x^\infty) = d^\infty$.

Familia de Penalidades descrita por Castillo R. y Gonzaga C.(ver [6])

En 1998, Castillo y Gonzaga [6] desarrollan dos familias de penalidades para ser utilizadas en métodos de Lagrangeano aumentado, con la novedad de que la derivada en la dirección +1 es finita e introduce un relevante y conveniente cambio de variables (shift) en ambas clases de penalidades, obteniendo que la derivada en el origen es igual a $\mu > 0$, estos cambios en las penalidades condujeron a distancias de Bregman vía dualidad de Fenchel y desarrollaron un método de Lagrangeano aumentado para cada una de ellas, cuya convergencia, para la sucesión dual, es garantizada por la teoría existente para distancias de Bregman, dando una prueba en el sentido ergódico para la convergencia de la sucesión en el primal. Castillo y Gonzaga desarrollan con estas familias un método de multiplicadores basado en shifts. Una breve descripción de estos métodos es la siguiente:

Dado el programa primal convexo (P) y suponiendo válidas las hipótesis (A1) y (A2) mencionadas anteriormente definen las siguientes familias de penalidades, que llamaron AL1 y AL2.

Familia de penalidades AL1: Sea $\theta : \mathbb{R} \to \mathbb{R}$, convexa, creciente y tal que:

(i) θ es estrictamente convexa.

(ii) θ es diferenciable en $\mathbb{R}$.

(iii) $\theta'_\infty(1) = 1$ es decir, 0-coerciva por la derecha.

(iv) $\lim\limits_{t\to-\infty} \theta'(y) = 0$.

(v) θ sea acotada inferiormente.

(vi) $sup_{y>0}\{y - \theta(y)\} < +\infty$.

Esta es una penalidad tipo I, además le agrega un parámetro $\beta \geq 1$ que aumenta la penalidad cuando sea necesario.

Familia de penalidades AL2: Sea $\theta : \mathbb{R} \to (-\infty, b)$, para $0 < b \leq +\infty$, tal que:

(i) θ es estrictamente convexa en $(-\infty, b)$.

(ii) θ es diferenciable en $(-\infty, b)$.

(iii) $\theta'_\infty(1) = +\infty$.

(iv) $\lim_{t \to -\infty} \theta^{'}(y) = 0$.

(v) θ sea acotada inferiormente.

Para las del tipo AL1, define la función:

$$\mu_i \in (0, \beta) \to \tilde{y}_i(\mu_i) = (\theta')^{-1} \left(\frac{\mu_i}{\beta_i} \right),$$

para $i = 1 \dots m$, o bien, vía dualidad de Fenchel;

$$\theta^{'}(\tilde{y}_i) = \frac{\mu_i}{\beta_i},$$

para $i = 1 \dots m$, donde $\frac{\mu_i}{\beta_i} < 1$. Para las del tipo AL2 define;

$$\mu_i \in \mathbb{R}^m_{++} \to \tilde{y}_i(\mu_i) = (\theta')^{-1}(\mu_i),$$

para $i = 1 \dots m$, o bien, vía dualidad de Fenchel

$$\theta^{'}(\tilde{y}_i) = \mu_i,$$

para $i = 1 \dots m$.

A estos cambios de variables los denominan shifts y permiten que cada familia tenga derivada μ en el origen, que es una condición importante al mirar su conjugada θ^*. Luego definen dos penalidades, para las AL1, la penalidad:

$$P_{\beta,r}(y, \mu) = \sum_{i=1}^{m} \beta_i r \left[\theta \left(\frac{y_i}{r} + \tilde{y}_i(\mu_i) \right) - \theta(\tilde{y}_i(\mu_i)) \right],$$

y para AL2 la penalidad:

$$P_r(y,\mu)=\sum_{i=1}^{m} r\left[\theta\left(\frac{y_i}{r}+\tilde{y}_i(\mu_i)\right)-\theta(\tilde{y}_i(\mu_i))\right].$$

Buscando las conjugadas de tales penalidades obtiene las conjugadas P_1^* y P_2^* dadas por,

$$P_1^*(s)=r\sum_{i=1}^{m}\left[\beta_i\theta^*\left(\frac{s_i}{\beta_i}\right)-\beta_i\theta^*\left(\frac{\mu_i}{\beta_i}\right)-(\theta^*)'\left(\frac{\mu_i}{\beta_i}\right)(s_i-\mu_i)\right],$$

y

$$P_2^*(s)=r\sum_{i=1}^{m}\left[\theta^*\left(s_i\right)-\theta^*\left(\mu_i\right)-(\theta^*)'\left(\mu_i\right)(s_i-\mu_i)\right].$$

con estas, ofrecen los algoritmos siguientes:

Algoritmo para la familia AL1:

Dados $\beta^0\geq e$; $r^0=1$; $\mu^0\in(0,\frac{\beta^0}{2})$ e $\tilde{y}^0$ tal que $\theta^{'}(\tilde{y}^0)=\frac{\mu^0}{\beta^0}$, haga

$$x^{k+1}\in argmin_{x\in\mathbb{R}^n}\{f_o(x)+\sum_{i=1}^{m}\theta_i^k(f_i(x))\}, \tag{2.7.23}$$

con

$$\theta_i^k(y_i)=\beta_i^k r^k(\theta(\frac{y_i}{r^k}+\tilde{y}_i^k)-\theta(\tilde{y}_i^k))$$

$$\mu_i^{k+1}=\beta_i^k\theta^{'}\left(\frac{f_i(x^{k+1})}{r^k}+\tilde{y}_i^k\right)\quad i=1,\ldots,m.$$

$$r^{k+1}=\frac{r^k}{2}$$

$$\beta^{k+1}=\frac{\beta^k}{2}.$$

Algoritmo para la familia AL2:

Dados $r^0=1$; $\mu^0=\theta^{'}(0)$ e $\tilde{y}^0=0$

$$x^{k+1}\in argmin_{x\in\mathbb{R}^n}\{f_o(x)+\sum_{i=1}^{m}\theta_i^k(f_i(x))\}, \tag{2.7.24}$$

con

$$\theta_i^k(y_i) = r^k\left(\theta\left(\frac{y_i}{r^k} + \tilde{y}_i^k\right) - \theta(\tilde{y}_i^k)\right)$$

$$\tilde{y}_i^{k+1} = \tilde{y}_i^k + y_i^{k+1}$$

donde,

$$y_i^{k+1} = \frac{f_i(x^{k+1})}{r^k}$$

$$\mu_i^{k+1} = \theta(\tilde{y}_i^{k+1}) \quad i = 1, \ldots, m.$$

Escoger $r^{k+1} \leq r^k$.

Estas P_1^* y P_2^* generan dos casi-distancias,

$$D_{h_1}(s,\mu) = r\left(h_1(s) - h_1(\mu) - \nabla h_1(\mu)^t(s-\mu)\right),$$

para

$$h_1(s) = \sum_{i=1}^{m} \beta_i \theta^*\left(\frac{s_i}{\beta_i}\right)$$

y

$$D_{h_2}(s,\mu) = r\left(h_2(s) - h_2(\mu) - \nabla h_2(\mu)^t(s-\mu)\right),$$

para

$$h_2(s) = \sum_{i=1}^{m} \theta^*(s_i),$$

resultando ambas distancias de Bregman para dichas funciones.

Con estas ideas desarrollan dos métodos de multiplicadores cuya convergencia queda expresada en el teorema (5.21) del mencionado trabajo [6]. Concretamente el teorema es el siguiente:

***Teorema* 2.7.2.** *Supóngase que son satisfechas (A1) y (A2) con respecto al problema (P), considere las suceciones* $\{x^k\}$, $\{\mu^k\}$, $\{r^k\}$ *y* $\{\beta^k\}$ *generadas por los métodos de multiplicadores con shift definidos en los algoritmos (AL1) o (AL2), para (AL2)* *considere además la sucesión* $\{y_i^k\}$ *tal que* $\theta^{'}(\tilde{y}_i^k) = \mu_i^k$*, i=1,...,m. y considere la sucesión de medias aritméticas en el primal,*$\{\bar{x}^k\}$ *con* $\bar{x}^k = \frac{\sum_{j=1}^k x^j}{k}$*. Entonces:*

(i) $lim_{k\to+\infty} sup f_i(\bar{x}^k)\} \leq 0,$ *para* $i = 1,\ldots,m.$

(ii) $lim_{k\to+\infty} \mu_i^k f_i(x^k) = 0,$ *para* $i = 1,\ldots,m.$

(iii) Todo punto límite de la sucesión de medias primal, $\{\bar{x}^k\}$ *es solución óptima de problema* (P)*.(Convergencia ergódica).*

Es precisamente en esta dirección en que se desarrolla este trabajo. En el siguiente capítulo se obtiene una casi-distancia que generaliza también esta situación, pero en lugar de un shift variable, usamos un shift constante y se obtienen resultados de convergencia en el sentido de que todos los puntos límites de las sucesiones primal y dual son soluciones óptimas de los problemas (P) y (D) respectivamente, en lugar de la convergencia ergódica.

Familia de Penalidades descrita por Auslender, Teboulle, Ben-Tiba.(ver [1])

Alfred Auslender, Marc Teboulle y Sami Ben Tiba desarrollan un algoritmo de punto proximal y métodos de multiplicadores para resolver programas convexos donde remplazan el término cuadrático por un núcleo cuadrático de segundo orden definido en términos de una función convexa. Básicamente consideran la familia de funciones de penalidad, la cual denotan por Φ con las siguientes características:

$\varphi : \mathbb{R} \to \mathbb{R}$ convexa, propia, cerrada con $dom(\varphi) \subset [0,+\infty), \varphi$ es C^2 en $int(dom(\varphi)) = (0,+\infty)$, además φ es estrictamente convexa sobre su dominio efectivo, $\lim\limits_{t\to 0^+} \varphi'(t) = -\infty, \varphi(1) = \varphi^{'}(1) = 0,$ $\varphi''(1) > 0$ y consideran también dos subfamilias Φ_1 y Φ_2 definidas así:

$\Phi_1 = \{\varphi \in \Phi : \exists M > 0/\varphi''(t) \leq M$ para todo $t \geq 1\}.$

Esta será una de las condiciones resultante en las funciones conjugadas de las funciones de penalidad sugeridas en este trabajo para construir la penalidad generalizada en el primal.

$\Phi_2 = \{\varphi \in \Phi : \varphi''(1)(1 - 1/t) \leq \varphi'(t) \leq \varphi''(1)(t-1)$, para todo $t > 0\}$.

Las funciones $\varphi_1(t) = tlog(t) - t + 1$, con $dom\varphi = [0, +\infty)$, $\varphi_2(t) = -log(t) + t - 1$, con $dom\varphi = (0, +\infty)$ y $\varphi_3(t) = 2(t^{1/2} - 1)^2$ con $dom\varphi = [0, +\infty)$ pertenecen a ambas familias.

Vía dualidad de Fenchel obtienen para $\varphi \in \Phi$ su conjugada φ^* con las propiedades $\varphi^*_\infty(-1) = 0; \varphi^*_\infty(1) = +\infty$. φ^* es estrictamente creciente, diferenciable sobre $int(dom(\varphi^*))$ y creciente, $int(dom(\varphi^*)) = (-\infty, \eta)$ con $0 < \eta \leq +\infty$.

Desarrollan el método siguiente: Para $\varphi \in \Phi, \lambda > 0$ y $\mu > 0$ defínase:

$H(x, \mu, \lambda) = f_0(x) + \lambda^{-1} \sum_{i=1}^m \mu_i^2 \varphi^*(\lambda f_i(x)/\mu_i)$ para $\mu_0 \in \mathbb{R}^m_{++}$ y $\lambda^k \geq \lambda > 0$, y $\varepsilon_k \geq 0$ para todo $k \geq 1$ genérese las sucesiones $\{x^k, \mu^k\}$ siguientes:

$$x^{k+1} \simeq argmin\{H(x, \mu^{k-1}, \lambda^k) : x \in \mathbb{R}^n\}$$

$$\mu_i^k = \mu_i^{k-1} (\varphi^*)' (\lambda^k f_i(x^k)/\mu_i^{k-1},$$

para $i = 1 \ldots m$, donde $\simeq$ significa que dado $\varepsilon_k \geq 0$, x^k debe satisfacer que $L(x^k, \mu^k) \leq inf_x L(x, \mu^k) + \varepsilon_k$, probando la convergencia de tal método bajo las condiciones siguientes:

Si $\varphi \in \Phi_1$, se cumple la condición de Slater, con $0 < \lambda < \lambda^k \leq \bar{\lambda}$ para todo k y $\sum \varepsilon_k < +\infty$, entonces $\{x^k\}$ y $\{\mu^k\}$ son acotadas y todos los puntos límites son soluciones óptimas de los problemas primal y dual.

En esta última década también existe un crecimiento importante en el estudio de estos métodos sobre variedades de Riemann, en [30] se encuentran suficientes herramientas que justifican el desarrollo de dichas teorías y suficientes ejemplos prácticos que justifican como en el ambiente Riemanniano la convexidad desarrolla todo su potencial. En [31] se encuentra la generalización del método de punto proximal sobre variedades de Riemann, en [32] se generaliza el método de punto proximal para variedades de Riemann con distancias de Bregman y en [30] un suficiente desarrollo de la convexidad en el ambiente Riemanniano e incluso aspectos preliminares de la teoría de dualidad que relaciona los problemas primal-dual Riemannianos y teoremas del tipo Karush-Khun Tucker en estos ambientes. Todo esto nos lleva a considerar la posibilidad de llevar la casi-distancia generalizada propuesta en el capítulo siguiente a los ambientes Riemannianos, claro está, pasando primero por un desarrolllo de la teoría de dualidad de Fenchel sobre estos ambientes Riemannianos y sus implicaciones para poder conectar los metodos de Lagrangeano con los de punto proximal, como fue realizado en este trabajo para el caso Euclídeo. Este planteamiento no se desarrollará en este trabajo y son sugerencias para futuras investigaciones.

Capítulo 3

Una Casi-distancia Generalizada.

3.1. Una introducción a la casi-distancia generalizada propuesta.

Este capítulo presenta los aportes originales de este trabajo, en la sección 3.1 se hace una breve introducción a la casi-distancia generalizada que se propone. En la sección 3.2 se presenta la familia de funciones de penalidad usadas y se construye la casi-distancia generalizada via dualidad de Fenchel. Se observa cómo la distancia entrópica Kullback-Leibler es generalizada por la casi-distancia propuesta y prueba que ésta es en efecto una medida divergente dándose además su interpretación geométrica. En la sección 3.3 se presenta el método propuesto y finalmente los teoremas de convergencia en los espacios primal (P) y dual (D). En resumen, este capítulo contiene básicamente el contenido de [33].

Considere el problema de programación convexa dado por

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

donde $f_i : \mathbb{R}^n \to \mathbb{R}$, para $i = 0, 1, \ldots, m$ son funciones convexas, propias y cerradas. El problema convexo dual asociado puede ser escrito como

$$(D) \quad d^* = \inf\{-d(\mu) \ : \ \mu \geq 0\} \tag{3.1.1}$$

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

$$l(x, \mu) = f_0(x) + \sum_{i=1}^{m} \mu_i f_i(x) \tag{3.1.2}$$

es la función Lagrangeana.

Suponga las hipótesis siguientes:

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

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

Como se ha visto, los métodos de Lagrangeano aumentado pueden ser introducidos desde un punto de vista primal, ver [4], [16],[3] o desde un punto de vista dual, donde el método de multiplicadores es construido aplicando toería de dualidad de Fenchel al método de punto proximal, ver [19], [11], [7], [1].

En la sección 3.2, se estudiará una nueva casi-distancia para resolver el problema (D) la cual tiene interesantes propiedades y está dada por la expresión

$$d^p_{\theta^*}(s,\mu) = \sum_{i=1}^{m} \left[\frac{\mu_i^p}{c} \theta^* \left(\frac{cs_i}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c)(s_i - \mu_i) \right] \tag{3.1.3}$$

donde θ^* es la función conjugada de la función de penalidad θ la cual no pasa a través del origen y satisface que $\theta'(0) = \kappa, \kappa > 0$, $p \geq 0$ y c es tal que $(\theta^*)'(c) = \tilde{y} \in \mathbb{R}$. Para $\mu \in \mathbb{R}^m_{++}$ fijo, si se define $h_\mu : \mathbb{R}^m_+ \to \mathbb{R}$ por $h_\mu(s) = \sum_{i=1}^m \frac{\mu_i^p}{c} \theta^* \left(\frac{cs_i}{\mu_i} \right)$ se obtiene $d^p_{\theta^*}(s,\mu) = h_\mu(s) - h_\mu(\mu) - \nabla h_\mu(\mu)^t (s - \mu)$, es decir, $d^p_{\theta^*}$ puede ser vista como la diferencia entre $h_\mu(s)$ y su aproximación lineal en $s = \mu$.

Se observará que θ^* es una traslación de la función φ usada en φ-divergencias, en este caso, con punto mínimo en $(\kappa, \theta^*(\kappa))$ y $\theta^*(\kappa) \in \mathbb{R}$. En la sección 3.3 se mostrarán resultados en los espacios primal y dual basados en un algoritmo de punto proximal y un algoritmo de Lagrangeano aumentado. Se consigue la casi-distancia generalizada aplicando la teoría de conjugacidad a una función de penalidad con shift, aplicada a un algoritmo de Lagrangeano aumentado, así se considerará primero una familia de funciones de penalidad y se mostrará como la casi-distancia generalizada definida en (3.1.7) surge naturalmente.

3.2. Consiguiendo la casi-distancia generalizada

Para obtener la casi-distancia generalizada de forma natural, se estudia primero una familia conveniente de funciones de penalidad en el contexto de los métodos de multiplicadores.

3.2.1. Funciones de Penalidad

Sea θ una función estrictamente creciente, convexa y dos veces diferenciable definida en $(-\infty, b)$, $0 < b \leq +\infty$, tal que:

1a) $0 < \theta'(0) = \kappa < +\infty$.

2a) $\lim_{t \to b} \theta'(t) = +\infty$.

3a) $\lim_{t \to -\infty} \theta'(t) = 0$.

4a) $\theta''(t) \geq \frac{1}{M}, \quad \forall t \in [0, b]$ y para algún $M > 0$.

La condición 1a) ha sido considerada en métodos con distancias de Bregman, ver [14], [6], pero no en casi-distancias, ver [11], [3], [16], [1], en este trabajo la función de penalidad no necesita pasar por el origen con pendiente uno. Asociada con la función θ, se considerará y se denotará por θ^* a su función conjugada, ver [17], la cual satisface las siguientes propiedades:

1b) θ^* es una función estrictamente convexa y diferenciable sobre $(0, +\infty)$.

2b) θ^* es decreciente en $(0, \kappa)$ y creciente en $(\kappa, +\infty)$ con $\theta^*(\kappa) \in \mathbb{R}$.

3b) $\lim_{t \to 0^+} (\theta^*)'(t) = -\infty$ y $\lim_{t \to +\infty} (\theta^*)'(t) = +\infty$.

4b) $(\theta^*)''(t) \leq M$ para todo $t \geq \kappa = \theta'(0)$.

La condición 2b) con $\kappa = 1$ es usada en el contexto de todas las casi-distancias conocidas actualmente asociadas a φ-divergencias.

Ejemplo:

Para $\theta(t) = e^{t+1}$, se tiene que $\theta(0) = e$, $\theta'(0) = e = \kappa$, $(\theta^*)'(e) = 0$ y $\theta^*(e) = -e$. Así, θ^* alcanza su mínimo en e cuya imagen es negativa $\theta^*(e) = -e$. La siguiente figura muestra las gráficas de θ y θ^*.

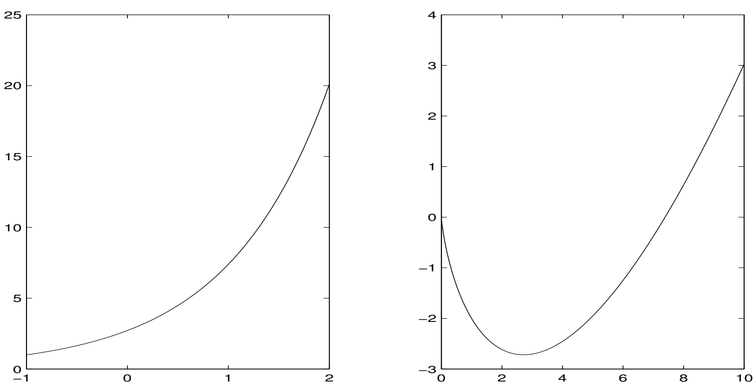

Fig.3.1 Graficas de θ and θ^*

3.2.2. Shift en funciones de penalidad

A diferencia de [14] y [6], en este trabajo se usa un shift constante para definir una función de penalidad generalizada la cual será utilizada en la sección 3.3, cuando se desarrolle el algoritmo en el contexto de métodos de multiplicadores.

Escójase $\tilde{y} \in \mathbb{R}$ tal que $\theta'(\tilde{y}) = c$ con $c \in \mathbb{R}_{++}$. Por el corolario 23.5.1 en [17] se tiene que:

$$(\theta')^{-1} = (\theta^*)' \tag{3.2.1}$$

y así

$$\theta'(\tilde{y}) = c \Leftrightarrow \tilde{y} = (\theta^*)'(c). \tag{3.2.2}$$

Dado $p \geq 0$, $r \in (0, 1]$, defínase la función de penalidad generalizada con shift P_p

como sigue:

$$y \in \mathbb{R}^m, \mu \in \mathbb{R}^m_{++} \mapsto P_p(y, \mu, r, c) = \sum_{i=1}^{m} P_{p,i}(y_i, \mu_i, r, c) \tag{3.2.3}$$

donde

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

Observe que $P_{p,i}(0, \mu_i, r, c) = 0$ para $i = 1, \ldots, m$ y

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

donde $(P_{p,i})'_1 = \frac{\partial P_{p,i}}{\partial y_i}$, así por (3.2.2)

$$(P_{p,i})'_1(0, \mu_i, r, c) = \frac{\mu_i}{c}\theta'(\tilde{y}) = \mu_i, \text{ for } i = 1, \ldots, m. \tag{3.2.4}$$

Observación

Un interesante hecho en la función de penalidad P_p es que se considera por primera vez exponentes racionales en los vectores multiplicadores aunque se probará la convergencia para valores de $p \geq 2$.

Geométricamente, el shift es una traslación que satisface la ecuación (3.2.4) . Las condiciones $\theta(0) = 0$ y $\theta'(0) = 1$ son utilizadas en todos los métodos conocidos actualmente basados en casi-distancias, ver [11], [1], [3], [16], sin embargo, en este trabajo, han sido relajadas, (ver la condicion 1a)) para trabajar con una familia más amplia de funciones.

Otro aspecto relevante del shift en esta función de penalidad, es que ésta permite construir una casi-distancia generalizada en el espacio dual aplicando teoría de conjugacidad, lo cual es mostrado en la siguiente proposición:

***Proposición* 3.2.2.1.** *Sea θ una función de penalidad que satisface las condiciones 1a)-4a). Dado $p \geq 0$, $r \in (0,1]$, $\tilde{y} \in \mathbb{R}$, $c \in \mathbb{R}_{++}$, considere*

$$y \in \mathbb{R}^m, \mu \in \mathbb{R}^m_{++} \mapsto P_p(y,\mu,r,c) = \sum_{i=1}^{m} r\frac{\mu_i^p}{c}\left[\theta\left(\frac{y_i}{\mu_i^{p-1}r}+\tilde{y}\right)-\theta(\tilde{y})\right] \tag{3.2.5}$$

donde $\theta'(\tilde{y}) = c$. Entonces

$$P^*_{p,\mu,r,c}(s) = r\left(\sum_{i=1}^{m}\frac{\mu_i^p}{c}\theta^*\left(\frac{cs_i}{\mu_i}\right)-\frac{\mu_i^p}{c}\theta^*(c)-\mu_i^{p-1}(\theta^*)'(c)(s_i-\mu_i)\right). \tag{3.2.6}$$

Demostración:

Considere

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

donde θ satisface las condiciones 1a)-4a) and $\theta'(\tilde{y}) = c$.

Considere fijos r, c, μ_i, para $i = 1,\ldots,m$ y la proposición 1.3.1 en [13], se tiene

$$\begin{aligned} P^*_{p,\mu_i,r,c}(s_i) &= \left[\tfrac{r}{c}\mu_i^p\theta\left(\tfrac{y_i}{\mu_i^{p-1}r}+\tilde{y}\right)-\tfrac{r}{c}\mu_i^p\theta(\tilde{y})\right]^* \quad \text{para } i=1,...,m \\ &= \left[\tfrac{r}{c}\mu_i^p\theta\left(\tfrac{y_i}{\mu_i^{p-1}r}+\tilde{y}\right)\right]^* + \tfrac{r}{c}\mu_i^p\theta(\tilde{y}) \\ &= \tfrac{r}{c}\mu_i^p\theta^*\left(\tfrac{cs_i}{\mu_i}\right) - r\mu_i^{p-1}s_i\tilde{y} + \tfrac{r}{c}\mu_i^p\theta(\tilde{y}). \end{aligned}$$

En efecto, considérese

$$f(y_i) = \theta(y_i+\tilde{y}) \Rightarrow f^*(s_i) = \theta(s_i) - s_i\tilde{y}$$

$$h(y_i) = \frac{r}{c}\mu_i^p f\left(\frac{y_i}{\mu_i^{p-1}r}\right) \Rightarrow h^*(s_i) = \frac{r}{c}\mu_i^p f^*\left(\frac{\mu_i^{p-1}rs_i}{\frac{r}{c}\mu_i^p r}\right)$$

Así,

$$\begin{aligned} h^*(s_i) &= \tfrac{r}{c}\mu_i^p f^*\left(\tfrac{cs_i}{\mu_i}\right) \quad \text{para } i = 1, ..., m \\ &= \tfrac{r}{c}\mu_i^p \theta^*\left(\tfrac{cs_i}{\mu_i}\right) - \tfrac{r}{c}\mu_i^p \tfrac{cs_i\tilde{y}}{\mu_i} \\ &= \tfrac{r}{c}\mu_i^p \theta^*\left(\tfrac{cs_i}{\mu_i}\right) - r\mu_i^{p-1} s_i\tilde{y}. \end{aligned}$$

Por el Teorema 25.3 en [17], si $\tilde{y} \in dom\,\theta$ y $c \in dom\,\theta^*$, se tiene que: $\theta(\tilde{y}) + \theta^*(c) = c\tilde{y}$. Usando (3.2.2) se tiene que para $i = 1, ..., m$,

$$\begin{aligned} P^*_{p,\mu_i,r,c}(s_i) &= \tfrac{r}{c}\mu_i^p \theta^*\left(\tfrac{cs_i}{\mu_i}\right) - \mu_i^{p-1} r s_i \tilde{y} + \tfrac{r}{c}\mu_i^p[c\tilde{y} - \theta^*(c)] \\ &= \tfrac{r}{c}\mu_i^p \theta^*\left(\tfrac{cs_i}{\mu_i}\right) - \tfrac{r}{c}\mu_i^p \theta^*(c) - \mu_i^{p-1} r\tilde{y}(s_i - \mu_i) \\ &= \tfrac{r}{c}\mu_i^p \theta^*\left(\tfrac{cs_i}{\mu_i}\right) - \tfrac{r}{c}\mu_i^p \theta^*(c) - \mu_i^{p-1} r(\theta^*)'(c)(s_i - \mu_i). \end{aligned}$$

Finalmente,

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

3.2.3. La casi-distancia generalizada

De acuerdo con la definición 2.1 en [12], dado $S \subset \mathbb{R}^n$; $\quad d : S \times S \to \mathbb{R}$ es llamada una medida divergente en S si y sólo si:

i) $d(x, y) \geq 0, \quad \forall x, y \in S$.

ii) If $\{x^k\} \subset S, x \in S$, entonces $\lim\limits_{k\to+\infty} d(x, x^k) = 0 \;\Leftrightarrow\; \lim\limits_{k\to+\infty} x^k = x$.

iii) El conjunto parcial de sub-nivel $\Gamma_1(y, \nu) = \{x \in S : d(x, y) \leq \nu\}$ es acotado $\forall y \in S \quad$ y $\quad \forall \nu > 0$.

iv) El conjunto parcial de sub-nivel $\Gamma_2(x,\nu) = \{y \in S : d(x,y) \leq \nu\}$ es acotado $\forall x \in S$ y $\forall \nu > 0$.

Usando (3.2.6), defínase la casi-distancia como:

$$d^p_{\theta^*}(x,y) = \sum_{i=1}^{m} \left[\frac{y_i^p}{c} \theta^* \left(\frac{cx_i}{y_i} \right) - \frac{y_i^p}{c} \theta^*(c) - y_i^{p-1} (\theta^*)'(c)(x_i - y_i) \right]. \tag{3.2.7}$$

En la siguiente proposición se probará que (3.2.7) es en efecto una medida divergente, pero antes obsérvense algunos detalles importantes. Note que $d^p_{\theta^*}(\cdot, y)$ es una función estrictamente convexa porque θ^* lo es.

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

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

la cual es usada en métodos de punto proximal con φ-divergencias, ver [12], [16].

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

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

los núcleos homogéneos de segundo orden usados en [1] y en [3] para un caso específico.

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

Para $m = 1$ y $\theta(t) = e^t$, se tiene que $\theta^*(s) = s\ln(s) - s, (\theta^*)'(s) = \ln(s)$ y $d^p_{\theta^*}(s,\mu) = \mu^{p-1} \left(s\ln(\frac{s}{\mu}) - s + \mu \right)$. Esta casi-distancia puede ser considerada una generalización de la distancia entrópica Kullback-Leibler. La misma expresión puede ser obtenida de $\theta(t) = e^t - 1$ o de $\theta(t) = e^{t+K}$ con $K \in R$.

Probemos ahora que la casi-distancia considerada en este trabajo es una medida divergente.

***Proposición* 3.2.3.1.** *Si θ^* satisface las condiciones 1b)-4b), $p \geq 0$, $r \in (0,1]$ y $c > 0$ entonces*

$$d^p_{\theta^*}(x,y) = \sum_{i=1}^{m} \left[\frac{y_i^p}{c} \theta^* \left(\frac{cx_i}{y_i} \right) - \frac{y_i^p}{c} \theta^*(c) - y_i^{p-1} (\theta^*)'(c)(x_i - y_i) \right]$$

es una medida divergente en R^n_{++}.

Demostración:

1) Ya que θ^* es convexa, $\forall\, z, w,\quad \theta^*(z) \geq \theta^*(w) + \theta^{*\prime}(w)(z-w)$, en particular para $w = c$ y $z = \frac{cx_i}{y_i}$, $\quad i = 1, ..., m$ con $x_i \geq 0, y_i > 0$ y dom $\theta^* = (0, +\infty)$,se tiene

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

y para todo $i = 1, ..., m$ y $p \geq 0$

$$\frac{y_i^p}{c}\theta^*\left(\frac{cx_i}{y_i}\right) - \frac{y_i^p}{c}\theta^*(c) - y_i^{p-1}(\theta^*)'(c)(x_i - y_i) \geq 0 \qquad (3.2.8)$$

así,

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

esto es, $d^p_{\theta^*}(x, y) \geq 0$.

2) Sea $\{x^k\}$ un sucesión en $(0, +\infty)$. Probemos que

$$\lim_{k\to+\infty} d^p_{\theta^*}(x, x^k) = 0 \;\Leftrightarrow \lim_{k\to+\infty} x^k = x.$$

Obsérvese que:

i) θ^* decreciente en $(0, \kappa)$ y creciente en $(\kappa, +\infty)$.

ii) $(\theta^*)'(\kappa) = 0$.

iii) $\lim\limits_{t\to+\infty} \theta^*(t) = +\infty$.

Así,

$$\lim_{k\to+\infty} d^p_{\theta^*}(x, x^k) = 0$$

$$\Leftrightarrow \lim_{k\to+\infty}\sum_{i=1}^{m}\left[\frac{(x_i^k)^p}{c}\theta^*\left(\frac{cx_i}{x_i^k}\right) - \frac{(x_i^k)^p}{c}\theta^*(c) - (x_i^k)^{p-1}\theta^{*\prime}(c)(x_i - x_i^k)\right] = 0,$$

$$\Leftrightarrow \lim_{k\to+\infty}\left[\frac{(x_i^k)^p}{c}\theta^*\left(\frac{cx_i}{x_i^k}\right) - \frac{(x_i^k)^p}{c}\theta^*(c) - (x_i^k)^{p-1}\theta^{*\prime}(c)(x_i - x_i^k)\right] = 0,\ i = 1, ..., m$$

$$\Leftrightarrow \lim_{k\to+\infty} \frac{(x_i^k)^p}{c}\theta^*\left[\theta^*\left(\frac{cx_i}{x_i^k}\right)-\theta^*(c)\right]-\left[(x_i^k)^{p-1}\theta^{*\prime}(c)(x_i-x_i^k)\right]=0,\ i=1,...,m$$
$$\Rightarrow \lim_{k\to+\infty} \frac{(x_i^k)^p}{c}\left[(\theta^*)'(\epsilon)\left(\frac{cx_i}{x_i^k}-c\right)\right]-\left[(x_i^k)^{p-1}\theta^{*\prime}(c)(x_i-x_i^k)\right]=0,\quad \frac{cx_i}{x_i^k}<\epsilon<c$$
$$\Rightarrow \lim_{k\to+\infty} (x_i^k)^{p-1}\left[(\theta^*)'(\epsilon)\left(x_i-x_i^k\right)\right]-(x_i^k)^{p-1}\left[\theta^{*\prime}(c)(x_i-x_i^k)\right]=0,\quad \frac{cx_i}{x_i^k}<\epsilon<c$$
$$\Rightarrow \lim_{k\to+\infty} x_i^k=x_i,\ i=1,...,m$$
$$\Rightarrow \lim_{k\to+\infty} x^k=x,\ i=1,...,m.$$

El recíproco es inmediato.

3) El conjunto de sub-nivel $\Gamma_1(y,\upsilon)=\{x\in(0,+\infty)/d^p_{\theta^*}(x,y)\leq\upsilon\}$ es acotado para todo $y\in(0,+\infty)$ y para todo $\upsilon\in(0,+\infty)$.

En efecto, supóngase que para algún $y\in(0,+\infty)$ y para algún $\upsilon\in(0,+\infty)$, $\Gamma(y,\upsilon)$ es no acotado.

Entonces existe $\{x^k\}$ en $(0,+\infty)$ / $x_k\to+\infty$, pero $d^p_{\theta^*}(x,y)\leq\upsilon$.

Por la Proposición 3.2.3.1, se tiene que:

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

Y para $i=1,\ldots,m$

$$0\leq\tfrac{y_i^p}{c}\theta^*\left(\tfrac{cx_i^k}{y_i}\right)-\tfrac{y_i^p}{c}\theta^*(c)-y_i^{p-1}(\theta^*)'(c)(x_i^k-y_i)\leq\upsilon,\quad \text{así}$$
$$0\leq\tfrac{y_i^p}{c}\theta^*\left(\tfrac{cx_i^k}{y_i}\right)-y_i^{p-1}(\theta^*)'(c)x_i^k\leq\upsilon+\tfrac{y_i^p}{c}\theta^*(c)-y_i^{p-1}(\theta^*)'(c)y_i,\quad \text{y}$$
$$0\leq\tfrac{y_i^p}{c}\theta^*\left(\tfrac{cx_i^k}{y_i}\right)-y_i^{p-1}(\theta^*)'(c)x_i^k\leq\upsilon+\tfrac{y_i^p}{c}\theta^*(c)-y_i^{p}(\theta^*)'(c).$$

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

$$0\leq\frac{y_i^p}{c}\theta^*\left(\frac{cx_i^k}{y_i}\right)-y_i^{p-1}(\theta^*)'(c)x_i^k\leq M_i\quad\text{y}$$

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

Por otro lado,

$$\lim_{k\to+\infty}\frac{y_i^p}{c}\left(\frac{\theta^*(\frac{cx_i^k}{y_i})}{x_i^k}\right)=\lim_{k\to+\infty}\frac{y_i^p}{c}\left(\frac{\theta^*(\frac{cx_i^k}{y_i})}{x_i^k}\right)=\frac{y_i^p}{c}\lim_{k\to+\infty}(\theta^*)'\left(\frac{cx_i^k}{y_i}\right)\frac{c}{y_i}$$
$$=y_i^{p-1}\cdot(+\infty)=+\infty.$$

Por tanto : $\frac{M_i}{x_i^k}+y_i^{p-1}(\theta^*)'(c)\to y_i^{p-1}(\theta^*)'(c)$, consiguiendo $+\infty\leq y_i^{p-1}(\theta^*)'(c)$, lo cual es una contradicción.

4) Se probará que $\Gamma_2(x,\upsilon)=\{y\in(0,+\infty)\quad/\quad d_{\theta^*}^p(x,y)\leq\upsilon\}$ es acotado para todo $x\in(0,+\infty)$ y para todo $\upsilon\in(0,+\infty)$.

Supóngase que existe $\{y^k\}\subset\Gamma_2(x,\upsilon)$ tal que para algún $j;\{y_j^k\}$ no es acotado y sin pérdida de generalidad $y_j^{k+1}>y_j^k$, para todo k. Para x fijo, ya que $\frac{x_j}{y_j^k}\to 0$ si $k\to+\infty$, entonces para k suficientemente grande: $\frac{x_j}{y_j^k}<1$ implicando que $\frac{cx_j}{y_j^k}<c$ y así,

$$\begin{aligned}
&\upsilon\geq\frac{(y_j^k)^p}{c}\theta^*\left(\frac{cx_j}{y_j^k}\right)-\frac{(y_j^k)^p}{c}\theta^*(c)-(y_j^k)^{p-1}(\theta^*)'(c)x_j+(y_j^k)^{p-1}(\theta^*)'(c)(y_j^k),\\
&\geq\frac{(y_j^k)^p}{c}(\theta^*)'(\xi)\,c\left[\frac{x_j}{y_j^k}-1\right]-(y_j^k)^p(\theta^*)'(c)\left(\frac{x_j}{y_j^k}\right)+(y_j^k)^p(\theta^*)'(c),\qquad\frac{cx_j}{y_j^k}<\xi<c.\\
&=(y_j^k)^p(\theta^*)'(\xi)\left[\frac{x_j}{y_j^k}-1\right]-(y_j^k)^p(\theta^*)'(c)\left[\frac{x_j}{y_j^k}-1\right],\\
&=(y_j^k)^p[(\theta^*)'(\xi)-(\theta^*)'(c)]\left[\frac{x_j}{y_j^k}-1\right],\\
&=(y_j^k)^p(\theta^*)''(\bar{\xi})(\xi-c)\left[\frac{x_j}{{y_j}^k}-1\right]\to+\infty\qquad(\xi<\bar{\xi}<c\ )
\end{aligned}$$

lo cual es una contradicción.

Finalmente $d_{\theta^*}^p(x,y)$ es una medida divergente.

3.2.4. Interpretación Geométrica.

Considere la casi-distancia generalizada definida en (3.2.7) y dado $\mu\in\mathbb{R}_{++}^m$, defínase la función convexa

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

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

$\nabla h_\mu(s) = \left(\mu_1^{p-1}(\theta^*)'\left(\frac{cs_1}{\mu_1}\right), \ldots, \mu_m^{p-1}(\theta^*)'\left(\frac{cs_m}{\mu_m}\right)\right)^t$

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

Se tiene entonces que

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

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

Esto muestra una interesante relación entre la casi-distancia generalizada y las distancias de Bregman, esto es, para cada $\mu > 0$, $h_\mu(s) = \sum_{i=1}^{m} \frac{\mu_i^p}{c}\theta^*\left(\frac{cs_i}{\mu_i}\right)$ es una función estrictamente convexa por causa de θ^*. Para cada $\mu > 0$, ésta genera una distancia de Bregman, ver [14], así se tiene que para cada $\mu > 0$, la casi-distancia generalizada puede escribirse como; $D_{h_\mu} : \mathbb{R}_+^m \times \mathbb{R}_{++}^m \to \mathbb{R}$ donde

$$D_{h_\mu}(s,\mu) = h_\mu(s) - h_\mu(\mu) - \nabla h_\mu(\mu)^t(s-\mu) = d_{\theta^*}^p(s,\mu). \tag{3.2.9}$$

Si se considera una sucesión $\{\mu^k\}$ en (3.2.9) se obtiene una sucesión de distancias de Bregman, tal que la casi-distancia generalizada en $\mu = \mu^k$ coincide con $D_{h_{\mu^k}}(x,y)$ en $y = \mu^k$. Esto sugiere la posibilidad de estudiar métodos de punto proximal con una sucesión de distancias de Bregman que dependan de un parámetro, o variando la función de Bregman inducida en cada iteración. Esto no será considerado en esta tesis.

La figura 2, muestra el gráfico de $h_\mu(s)$ y su aproximación lineal en μ para μ=2, $\frac{3}{2}$, 1, $c = \frac{1}{2}$, $p = \frac{5}{2}$ y θ^* como en el ejemplo 1. Del lado derecho, el gráfico de $d_{\theta^*}^p(s,\mu)$ para los mismos valores.

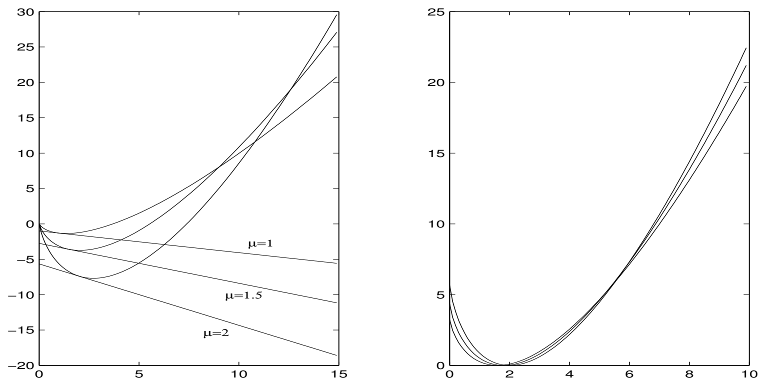

Fig.3.2 Gráfica de $d^p_{\theta^*}(s,\mu)$ para algunos valores particulares de c, p y μ

3.3. Métodos y Teoremas.

3.3.1. Método Proximal.

Considere la casi-distancia generalizada aplicada directamente sobre el problema dual. El método proximal para resolver el problema (D) definido en (3.1.1) genera una sucesión $\{\mu^k\}$ tal que $\mu^0 \in \mathbb{R}^m_{++}$ y

$$\mu^{k+1} = argmin\{-d(\mu) + r^k d^p_{\theta^*}(\mu, \mu^k)\}, \tag{3.3.1}$$

donde $r^k \in [\widehat{r}, \overline{r}] \subset (0, +\infty)$ y θ^* como se definió en la sección (3.2.1).

Por la condición de optimalidad se tiene que

$$0 \in \partial \left[-d(\mu^{k+1}) + r^k d^p_{\theta^*}(\mu^{k+1}, \mu^k)\right]$$

o equivalentemente

$$-r^k \left((\mu_1^k)^{p-1}[\theta^{*\prime}\left(\frac{c\mu_1^{k+1}}{\mu_1^k}\right) - \theta^{*\prime}(c)], ..., (\mu_m^k)^{p-1}[\theta^{*\prime}\left(\frac{c\mu_m^{k+1}}{\mu_m^k}\right) - \theta^{*\prime}(c)] \right) \in \partial(-d(\mu^{k+1})). \tag{3.3.2}$$

3.3.2. Resultados de convergencia.

***Proposición* 3.3.2.1.** $\{-d(\mu^k)\}$ *es una sucesión no creciente y convergente.*

Demostración. Por la condición de optimalidad en (3.3.1) se tiene $\mu^{k+1} = argmin_{\mu>0}\{-d(\mu)+r^k d^p_{\theta*}(\mu,\mu^k)\}$, así $-d(\mu^{k+1})+r^k d^p_{\theta*}(\mu^{k+1},\mu^k) \leq -d(\mu^k)+ r^k d^p_{\theta*}(\mu^k,\mu^k) = -d(\mu^k)$, esto es $-d(\mu^{k+1}) \leq -d(\mu^k)$. Por teoría de dualidad, f^* es una cota inferior para $\{-d(\mu^k)\}$, luego $\{-d(\mu^k)\}$ es convergente. □

***Proposición* 3.3.2.2.** *La sucesión* $\{\mu^k\}$ *generada por (3.3.1) es acotada.*

Demostración. Por (H2) y ya que $-d$ es una función propia convexa, el conjunto de nivel $\Lambda = \{\mu \in \mathbb{R}^m_+ \,/ -d(\mu) \leq -d(\mu_0)\}$ es compacto y por la proposición 3.3.2.1, $\mu^k \in \Lambda$ para todo k, así μ^k es acotada.

□

En la próxima sección, en base a la Proposición 3.2.2.1, un método de multiplicadores primal asociado al proximal puede ser construido usando las funciones de penalidad generalizada definida en [10].

3.3.3. Algoritmo de Lagrangeano Aumentado.

Considere el problema (P) con hipótesis (H1) y (H2) y la θ-funciones cumpliendo las condiciones 1a)-4a) en la sección (3.2.1).

Dando $p \geq 0$, $r \in (0,1]$, $\tilde{y} \in \mathbb{R}$, $c \in \mathbb{R}_{++}$ con $\theta'(\tilde{y}) = c$, la función Lagrangeana aumentada está dada por

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

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

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

El método de multiplicadores asociado al método de punto proximal está dado por:

$$x^{k+1}\in \text{argmin}\{f_0(x)+\sum_{i=1}^{m}P_{p,i}(f_i(x),\mu_i^k,r^k,c)\}, \tag{3.3.3}$$

$$\mu_i^{k+1}=\frac{\mu_i^k}{c}\theta'\left(\frac{f_i(x^{k+1})}{(\mu_i^k)^{p-1}r^k}+\tilde{y}\right) \quad \text{para } i=1,\cdots,m \text{ y } r^k\in(0,1]. \tag{3.3.4}$$

Observe 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.

La próxima proposición muestra que la sucesión definida en (3.3.4) y (3.3.1) son la misma.

***Proposición* 3.3.3.1.** *Sea $\{\widehat{\mu}^k\}$ una sucesión generada por (3.3.1) que resuelve el problema dual (D) y sean $\{x^k\}$ y $\{\mu^k\}$ las sucesiones generadas por (3.3.3) y (3.3.4) que resuelven el problema primal (P). Si $\mu^0=\widehat{\mu}^0$, entonces $\forall k>0:\mu^k=\widehat{\mu}^k$.*

Supongamos por inducción que $\mu^k=\widehat{\mu}^k$.

Para todo $\mu\in\mathbb{R}_+^m$ se tiene que:

$$\begin{aligned}
&d(\mu)=inf_x\{f_0(x)+\sum_{i=1}^{m}\mu_i f_i(x)\}\\
&\le f_0(x^{k+1})+\sum_{i=1}^{m}\mu_i f_i(x^{k+1})\\
&=f_0(x^{k+1})+\sum_{i=1}^{m}\mu_i^{k+1}f_i(x^{k+1})+\sum_{i=1}^{m}(\mu_i-\mu_i^{k+1})f_i(x^{k+1})\\
&=d(\mu^{k+1})+\sum_{i=1}^{m}(\mu_i-\mu_i^{k+1})f_i(x^{k+1}).
\end{aligned}$$

Así, por ser d cóncava, se tiene:

$$f(x^{k+1})\in\partial d(\mu^{k+1})$$

donde $f(.) = (f_i(.), \ldots, f_m(.))$ y $\partial d(\mu)$, es el conjunto de subgradientes en μ de la función cóncava d

Pero

$$\mu_i^{k+1} = \frac{\mu_i^k}{c}\theta'\left(\frac{f_i(x^{k+1})}{(\mu_i^k)^{p-1}r} + \tilde{y}\right),$$

$$\frac{c\mu_i^{k+1}}{\mu_i^k} = \theta'\left(\frac{f_i(x^{k+1})}{(\mu_i^k)^{p-1}r} + \tilde{y}\right),$$

$$(\theta^{'})^{-1}\left(\frac{c\mu_i^{k+1}}{\mu_i^k}\right) - \tilde{y} = \frac{f_i(x^{k+1})}{(\mu_i^k)^{p-1}r},$$

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

$$\mu^{k+1} \in argmin\{-d(\mu) + r^k d_{\theta^*}^p(\mu, \hat{\mu}^k)\}.$$

3.3.4. Resultados de Convergencia.

Esta sección está inspirada en los teoremas de convergencia en [3] y [16]. Para simplificar, la ecuación (3.2.7) se escribirá de la forma siguiente:

$$d_{\theta^*}^p(s,\mu) = \sum_{i=1}^m d(s_i, \mu_i), \tag{3.3.5}$$

donde $d(s_i,\mu_i) = \frac{\mu_i^p}{c}\theta^*\left(\frac{cs_i}{\mu_i}\right) - \frac{\mu_i^p}{c}\theta^*(c) - \mu_i^{p-1}(\theta^*)'(c)(s_i - \mu_i)$ para $i = \ldots, m$ y así $d_1'(s_i,\mu_i) = \frac{\partial d(s_i,\mu_i)}{\partial s_i} = \mu_i^{p-1}\left[(\theta^*)'\left(\frac{cs_i}{\mu_i}\right) - (\theta^*)'(c)\right]$.

***Lema* 3.3.4.1.** *Sea s, μ números reales positivos tales que $s > \mu$ entonces*

$$d(s,\mu) \geq \frac{1}{2}\frac{[d_1'(s,\mu)]^2}{c\,M\mu^{p-2}},$$

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

Demostración. Considere la función cuadrática

$$q(t) = q(s) + (t-s)d_1'(s,\mu) + \frac{1}{2}(t-s)^2 cM\mu^{p-2},$$

así

$$\begin{aligned} q'(t) &= d_1'(s,\mu) + (t-s)cM\mu^{p-2} \ \text{ y} \\ q'(t) = 0 &\Leftrightarrow d_1'(s,\mu) + tcM\mu^{p-2} - scM\mu^{p-2} = 0 \\ &\Leftrightarrow t = \tfrac{-d_1'(s,\mu)}{cM\mu^{p-2}} + \tfrac{scM\mu^{p-2}}{cM\mu^{p-2}} = \tfrac{-d_1'(s,\mu)}{cM\mu^{p-2}} + s, \end{aligned}$$

luego

$$t^* = \frac{-d_1'(s,\mu)}{cM\mu^{p-2}} + s \tag{3.3.6}$$

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

Ya que $s > \mu > 0$

$$d_1'(s,\mu) = \mu^{p-1}\left[(\theta^*)'\left(\frac{cs}{\mu}\right) - (\theta^*)'(c)\right] > 0.$$

Así , por (3.3.6) $t^* < s$. Por tanto,

$$\begin{aligned} d_1'(s,\mu) &= d_1'(s,\mu) - d_1'(\mu,\mu) = (s-\mu)d_1''(\hat{\mu},\mu) \ \text{ para algún } \hat{\mu} \in [\mu, s] \\ &= (s-\mu)\mu^{p-2}(\theta^*)''(\tfrac{c\hat{\mu}}{\mu})c \le (s-\mu)\mu^{p-2}cM. \end{aligned}$$

Así $\frac{d_1'(s,\mu)}{cM\mu^{p-2}} \le s - \mu$ y

$$\mu \le s - \frac{d_1'(s,\mu)}{cM\mu^{p-2}} = t^*, \tag{3.3.7}$$

entonces $\mu \le t^* < s$.

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

$$\begin{aligned} d_1'(s,\mu) - d_1'(t,\mu) &= (s-t)d_1''(\bar{t},\mu) \text{ para algún } t \le \bar{t} \le s \\ &= (s-t)\mu^{p-2}(\theta^*)''(\tfrac{c\bar{t}}{\mu})c \\ &\le (s-t)\mu^{p-2}cM, \end{aligned}$$

así $d_1'(s,\mu) \le d_1'(t,\mu) + (s-t)\mu^{p-2}cM$ para todo $\mu \le t \le s$

y

$$d_1'(t,\mu) \geq d_1'(s,\mu) + (t-s)\mu^{p-2}cM = q'(t),$$

se tiene entonces

$$d_1'(t,\mu) \geq q'(t) \text{ para todo } \mu \leq t \leq s.$$

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

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

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

$$d(s,\mu) \geq q(s) - q(s) - (t^*-s)d_1'(s,\mu) - \frac{1}{2}(t^*-s)^2 cM\mu^{p-2}. \tag{3.3.8}$$

Por (3.3.7) $t^* - s = \frac{-d_1'(s,\mu)}{cM\mu^{p-2}}$,

reemplazando en (3.3.8) se tiene

$$\begin{aligned} d(s,\mu) &\geq \tfrac{[d_1'(s,\mu)]^2}{cM\mu^{p-2}} - \tfrac{1}{2}\tfrac{[d_1'(s,\mu)]^2}{cM\mu^{p-2}} \\ &= \tfrac{1}{2}\tfrac{[d_1'(s,\mu)]^2}{cM\mu^{p-2}}. \end{aligned} \tag{3.3.9}$$

Finalmente,

$$d(s,\mu) \geq \frac{1}{2}\frac{[d_1'(s,\mu)]^2}{cM\mu^{p-2}}.$$

□

La próxima proposición utiliza implícitamente una variante de la propiedad (D4) en [3], y es válida para $p \geq 2$.

***Proposición* 3.3.4.1.** *(Factibilidad asintótica) Considere* $p \geq 2$. *La sucesión* $\{[f_i(x^k)]_+\}$ *converge a 0 para todo* $i = 1, ..., m$ *donde* $[y]_+ = max\{0, y\}$.

Demostración. Supóngase por reducción al absurdo que $\{[f_i(x^k)]\} \nrightarrow 0$, entonces existe $\{l_k\}$ tal que $l_k \to +\infty$ con $l_k < l_{k+1}$ y $\exists \epsilon > 0$ tal que $f_{i_0}(x^{l_k}) > \epsilon$ para algún $i_0 \in \{i = 1, \ldots, m\}$.

Se tiene $d(\mu) = inf\{f_0(x) + \sum_{i=1}^{m} \mu_i f_i(x)\} \leq f_0(x^{k+1}) + \sum_{i=1}^{m} \mu_i f_i(x^{k+1})$

$= f_0(x^{k+1}) + \sum_{i=1}^{m} \mu_i^{k+1} f_i(x^{k+1}) + \sum_{i=1}^{m} (\mu_i - \mu_i^{k+1}) f_i(x^{k+1})$

$= d(\mu_i^{k+1}) + \sum_{i=1}^{m} (\mu_i - \mu_i^{k+1}) f_i(x^{k+1})$ para $i = 1, \ldots, m$.

Por (3.3.4),

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

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

$$r^k (\mu_i^k)^{p-1} \left[(\theta^*)' \left(\frac{c(\mu_i^{k+1})}{\mu_i^k} \right) - (\theta^*)'(c) \right] = f_i(x^{k+1}) \quad \text{para } i = 1, \ldots, m. \qquad (3.3.10)$$

Usando la notación en (3.3.5),

$$d(s_i, \mu_i) = \frac{\mu_i^p}{c} \theta^* \left(\frac{c s_i}{\mu_i} \right) - \frac{\mu_i^p}{c} \theta^*(c) - \mu_i^{p-1} (\theta^*)'(c)(s_i - \mu_i) \quad \text{para } i = 1, \ldots, m$$

y

$$d_1'(s_i, \mu_i) = \mu_i^{p-1} \left[(\theta^*)' \left(\frac{c s_i}{\mu_i} \right) - (\theta^*)'(c) \right] \quad \text{para } i = 1, \ldots, m. \qquad (3.3.11)$$

Por (3.3.10) y (3.3.11)

$$\frac{f_i(x^{k+1})}{r^k} = d_1'(\mu_i^{k+1}, \mu_i^k) \quad \text{para } i = 1, \ldots, m. \qquad (3.3.12)$$

Ya que $f_{i_0}(x^{l_{k+1}}) > \epsilon$ para $i_0 \in \{1, \ldots, m\}$, $r^k d_1'(\mu_{i_0}^{l_{k+1}}, \mu_{i_0}^{l_k}) > \epsilon > 0$. Esto implica $d_1'(\mu_{i_0}^{l_{k+1}}, \mu_{i_0}^{l_k}) > 0$ y así $\mu_{i_0}^{l_{k+1}} > \mu_{i_0}^{l_k}$ y se puede usar el lema 3.3.4.1.

Ya que $\mu^{k+1} = argmin\{-d(\mu) + d_{\theta^*}^p(\mu, \mu^k)\}$, se tiene que:

$$-d(\mu^{k+1}) + d_{\theta^*}^p(\mu^{k+1}, \mu^k) \leq -d(\mu^k) + d_{\theta^*}^p(\mu^k, \mu^k) = -d(\mu^k),$$

luego, $d(\mu_{k+1}) - d(\mu_k) \geq d^p_{\theta^*}(\mu^{k+1}, \mu^k) \geq 0$.

Entonces $d(\mu^{l_{k+1}}) - d(\mu^{l_k}) \geq d^p_{\theta^*}(\mu^{l_{k+1}}, \mu^{l_k}) = \sum_{i=1}^{m} d(\mu_i^{l_{k+1}}, \mu_i^{l_k}) \geq d(\mu_{i_0}^{l_{k+1}}, \mu_{i_0}^{l_k})$. Por la proposición 3.3.2.2, sea $\widehat{\mu}$ una cota superior de $\{\mu^k\}$ y por lema 3.3.4.1 se tiene

$$d(\mu^{l_{k+1}}) - d(\mu^{l_k}) \geq \frac{1}{2cM}\frac{[d_1'(\mu_{i_0}^{l_{k+1}}, \mu_{i_0}^{l_k})]^2}{(\mu_{i_0}^{l_k})^{p-2}} \geq \frac{\epsilon^2}{2cM\widehat{\mu}_{i_0}^{p-2}} = \delta > 0$$

para $p \geq 2$.

Entonces $d(\mu^{l_{k+1}}) - d(\mu^{l_k}) \geq \delta > 0$ así tomando límite cuando $k \to \infty$ se tiene que 0 es mayor que un valor positivo , lo cual es una contradicción.

Finalmente, $\lim\limits_{k\to+\infty} [f_i(x^k)]_+ = 0$. □

La proxima proposición utiliza la siguiente afirmación:

Afirmación 1: Para todo t, $t\frac{\theta'(t+\widetilde{y})}{c} \geq t$.

Demostración. Si $t > 0 \Rightarrow t + \widetilde{y} > \widetilde{y} \Rightarrow \theta'(t + \widetilde{y}) > \theta'(\widetilde{y}) = c$

$\Rightarrow \frac{\theta'(t+\widetilde{y})}{c} > 1 \Rightarrow t\frac{\theta'(t+\widetilde{y})}{c} > t$.

If $t < 0 \Rightarrow t + \widetilde{y} < \widetilde{y} \Rightarrow \theta'(t + \widetilde{y}) < \theta'(\widetilde{y}) = c$

$\Rightarrow \frac{\theta'(t+\widetilde{y})}{c} < 1 \Rightarrow t\frac{\theta'(t+\widetilde{y})}{c} > t$. □

***Proposición* 3.3.4.2.** *Sean* $\{x^k\}$ *y* $\{\mu^k\}$ *las sucesiones generadas por (3.3.3) y (3.3.4), entonces* $\lim_{k\to+\infty} \mu_i^k f_i(x^k) = 0$ *para* $i = 1, \cdots, m$.

Demostración. Por reducción al absurdo, supóngase que existe $i_0 \in \{1, ..., m\}, \epsilon > 0$ y un conjunto infinito de índices k_j tal que

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

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

$$0 < \mu^k \leq \widehat{\mu} \quad \text{para todo } k, \tag{3.3.13}$$

entonces

$$|\mu_{i_0}^{k_j+1} f_{i_0}(x^{k_j+1})| \geq \epsilon \Rightarrow |f_{i_0}(x^{k_j+1})| \geq \frac{\epsilon}{\widehat{\mu}_{i_0}}.$$

Pero $\{[f_i(x^k)]_+\}$ converge a 0, para todo $i = 1, ..., m$ entonces $f_{i_0}(x^{k_j+1}) \geq \frac{\epsilon}{\widehat{\mu}_{i_0}}$ es cierto sólo para un conjunto finito de índices k_j, así se puede considerar sin pérdida de generalidad que

$$f_{i_0}(x^{k_j+1}) \leq -\frac{\epsilon}{\widehat{\mu}_{i_0}}, \tag{3.3.14}$$

para todo j.

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

$$\sum_{i=1}^{m} f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \leq d(\mu^{k_j+1}) - d(\mu^{k_j}). \tag{3.3.15}$$

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

$$\begin{aligned} f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) &= \mu_i^{k_j} f_i(x^{k_j+1}) \left(\frac{\mu_i^{k_j+1}}{\mu_i^{k_j}} - 1\right), \\ &= \mu_i^{k_j} f_i(x^{k_j+1}) \left(\frac{1}{c}\theta'\left(\frac{f_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1}r^k} + \widetilde{y}\right) - 1\right), &(3.3.16) \\ &= \frac{\mu_i^{k_j} f_i(x^{k_j+1})}{c}\theta'\left(\frac{f_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1}r^k} + \widetilde{y}\right) - \mu_i^{k_j} f_i(x^{k_j+1}). &(3.3.17) \end{aligned}$$

Ya que θ es una función estríctamente convexa, $\theta''(t) > 0$ para todo $t \in (-\infty, b)$ entonces θ' es creciente y así usando (3.3.17) y la afirmación 1

$$\begin{aligned} f_i(x^{k_j+1})(\mu_i^{k_j+1)} - \mu_i^{k_j}) &= \frac{r^k(\mu_i^{k_j})^p}{c}\left[\frac{f_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1}r^k}\theta'\left(\frac{f_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1}r^k} + \widetilde{y}\right)\right] - \mu_i^{k_j} f_i(x^{k_j+1}) \\ &\geq \mu_i^{k_j} f_i(x^{k_j+1}) - \mu_i^{k_j} f_i(x^{k_j+1}) \\ &= 0, \end{aligned}$$

entonces $\quad 0 \leq \sum_{i=1}^{m} f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \leq d(\mu^{k_j+1}) - d(\mu^{k_j}),$

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

y así

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

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

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

Así por (3.3.17),

$$\lim_{j\to+\infty} \frac{\mu_i^{k_j} f_i(x^{k_j+1})}{c} \left[\theta'\left(\frac{f_i(x^{k_j+1})}{(\mu_i^{k_j})^{p-1} r^k} + \widetilde{y}\right) - 1\right] = 0 \text{ para todo } i = 1, ..., m. \quad (3.3.18)$$

Por otro lado,

$$\frac{f_{i0}(x^{k_j+1})}{r^k(\mu_{i_0}^{k_j})^{p-1}} + \widetilde{y} \leq \frac{-\epsilon}{r^k(\mu_{i_0}^{k_j})^{p-1}\widehat{\mu_{i0}}} + \widetilde{y} < 0 + \widetilde{y},$$

y ya que θ' es creciente

$$\theta'\left(\frac{f_{i_0}(x^{k_j+1})}{r^k(\mu_{i_0}^{k_j})^{p-1}} + \widetilde{y}\right) \leq \theta'\left(\frac{-\epsilon}{r^k(\mu_{i_0}^{k_j})^{p-1}\widehat{\mu_{i_0}}} + \widetilde{y}\right) < \theta'(\widetilde{y}) = c,$$

así

$$\frac{1}{c}\theta'\left(\frac{f_{i_0}(x^{k_j+1})}{r^k(\mu_{i_0}^{k_j})^{p-1}} + \widetilde{y}\right) \leq \frac{1}{c}\theta'\left(\frac{-\epsilon}{r^k(\mu_{i_0}^{k_j})^{p-1}\widehat{\mu_{i_0}}} + \widetilde{y}\right) < 1. \quad (3.3.19)$$

Por (3.3.18) y (3.3.19)

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

Ya que

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

entonces $\lim_{j\to+\infty} \mu_{i_0}^{k_j+1} f_{i_0}(x^{k_j+1}) = 0$, lo cual es una contradicción.

Consecuentemente

$$\mu_i^k f_i(x^k) \to 0$$

para todo $i = 1, ..., m$ □

***Teorema* 3.3.4.1.** *La sucesión $\{f_0(x^k)\}$ converge a $\widehat{f}$, cada punto límite de la sucesión $\{x^k\}$ y $\{\mu^k\}$ generada por el método (3.3.3) y (3.3.4) (o (3.3.3) y (3.3.1)) son soluciones óptimas del problema (P) y (D) respectivamente.*

Demostración. Ya que $\{x^k\}$ asintóticamente factible, para todo $\epsilon > 0, f_0(x^k) \geq \widehat{f} - \epsilon$ para k suficientemenete grande.

De acuerdo a (H2), el conjunto de valores óptimos del problema (D) es no vacío y compacto, $\widehat{f} = \widehat{d}$ donde

$$\widehat{f} = \min\{f_0(x) : x \in R^n, f_i(x) \leq 0, i = 1, ..., m\}$$

y

$$\widehat{d} = \sup\{d(\mu) : \mu \in R_+^m\}.$$

Por tanto, para cada $\beta < \widehat{d}$ el conjunto de nivel $\{\mu \in R_+^m : d(\mu) \geq \beta\}$ es compacto.

Pero

$$\widehat{f} = \widehat{d} \geq d(\mu^k) = \inf_x\{l(x, \mu^k)\} = l(x^k, \mu^k) = f_0(x^k) + \sum_{i=1}^{m} \mu_i^k f_i(x^k) \tag{3.3.20}$$

y debido a la proposición 3.3.4.2 $\{\mu_i^k f_i(x^k)\}$ converge a 0 para todo $i = 1, ..., m$ así para todo $\epsilon > 0, \widehat{f} - \epsilon \leq f_0(x^k) \leq \widehat{f} - \sum_{i=1}^{m} \mu_i^k f_i(x^k) < \widehat{f} + \epsilon$ para k suficientemente grande; se tiene entonces

$$f_0(x^k) \to \widehat{f}. \tag{3.3.21}$$

Por la proposición 3.3.4.1, para todo $\epsilon > 0$ y para $i = 1, ..., m$, $f_i(x^k) \leq \epsilon$ para k suficientemente grande, entonces $f_0(x^k) \leq \widehat{f} + \epsilon$ y $f_i(x^k) \leq \epsilon$ para $i = 1, ..., m$ y k suficientemente grande. Debido a (H1) y el corolario 20 in [8] el conjunto

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

es compacto para cualquier $\alpha, \ \beta$, entonces $\{x^k\}$ es acotado y por la proposición 3.3.2.2 la sucesión $\{\mu^k\}$ también es acotada. Si $\bar{x}$ es un punto límite de $\{x^k\}$, entonces por la proposición 3.3.4.1 y (3.3.21), $\bar{x}$ es una solución óptima primal y si $\bar{\mu}$ es un punto límite de $\{\mu^k\}$, usando (3.3.20) y la proposición 3.3.4.2, $d(\bar{\mu}) = \widehat{f} = \widehat{d}$. Por tanto $\bar{\mu}$ es una solución óptima dual. $\square$

Capítulo 4

Conclusiones.

4.1. Observaciones finales

Primeramente se concluye que las casi-distancias asociadas a φ divergencias, a núcleos homogeneos de segundo orden y a dsitancias de Bregman son objetos matemáticos que guardan una estrecha relación entre si, ya que son casos particulares de una casi-distancia generalizada. Además, tal casi distancia tiene la particularidad de que representa un núcleo no cuadrático que proviene de una familia de funciónes de penalidad que no necesariamente pasan por el origen ni necesariamente tienen pendiente uno y la familia de funciones de penalidad considerada permite construir un método de multiplicadores que por primera vez considera potencias racionales p en el vector multiplicador.

Un problema abierto que se deriva de esta investigación es estudiar cual es el mejor valor de p que nos permite obtener la mejor rata de convergencia en el contexto de los métodos de multiplicadores, aunque se observó en pruebas computacionales que el número de iteraciones en el algoritmo principal decrece a medida que p crece, a pesar de presentar un mal comportamiento numérico.

Métodos de multiplicadores sin shift y para $p = 3$ fueron considerados en [5] y son un caso particular del estudio realizado en este trabajo.

Un analisis de convergencia para el caso $p = 1$ puede ser obtenido dando hipótesis y teoremas similares a los presentados en [11].

En [15] se consideraron los resultados de convergencia a lo que sería el caso $p = 0$ para una función particular de penalidad sin shift.

Otras consideraciones acerca de los valores de p, por ejemplo, $p < 0$, $0 < p < 1$ y $1 < p < 2$ permanecen aún como problemas abiertos.

Por otro lado el método de punto proximal desarrollado con la casi-dsitancia generalizada suguiere que se deben estudiar métodos de punto proximal asociados a familias de distancias de Bregman que dependan de un parámetro, como se observó que ocurre con la casi-distancia generalizada.

Finalmente, se podrían considerar shifts en cada iteración, como $\theta(\tilde{y}_i^k) = c_i^k$ y escoger $c_i^k = r^k \alpha^k$ como en [18] con la intención de relacionar ambas aproximaciones.

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. 645-668, 1999.

[2] Mokhtar S. Bazaraa and C.M. Shetty, " Nonlinear Pogramming, 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. 447-366, 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 funtions, with aplications to convex programming", Mathematics of Operations Researchs, 18, pp. 202-226, 1993.

[8] A. Fiacco and G. McCormick, Nonlinear programming: Sequential Unconstrained Minimization Techniques, Classics in applied Mathemathics, SIAM, Philadhelpia, 1990.

[9] A. Iusem, Metodos de ponto proximal em Otimizacao, 20 Coloquio Brasileiro de matemática," IMPA, R.J, Brazil, 1995.

[10] Censor, Y., Zenios, S. "The proximal minimizations algorithm with D-functions.", Journal of Optimization Theory and Applications 73 (1992) 451-464.

[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", Investigacion Operativa, 8 , pp. 11-50, 1997.

[13] Hiriart-Urruty and C. Lemarechal, " Convex Analysis and Minimization Algorithm I", Springer Verlag, New York, 1996.

[14] K. Kiwiel , " Proximal minimization methods with generalized Bregman funtions", SIAM J. on Control and Optimization, 35, 1142-1168, 19(4), 1997.

[15] L. Matioli, Tese de Doutorado, Univesidade Federal de Santa Catarina, Brazil, 2000.

[16] R. Poliak and M.Teboulle, "Nonlinear rescaling and proximal-like methods in convex programming, 76, pp. 697-739, 1997.

[17] R.T. Rockafellar, "Convex Analysis, Princeton University Press, Princeton, N,J, 1970.

[18] P.J. Silva, J. Eckstein and C. Humes, "Rescaling and Stepsize selection in proximal methods using separable generalized distances," SIAM Journal on Optimization 12(1), pp. 238-261, 2001.

[19] M. Teboulle, "Entropic proximal mappings with applied to nonlinear programming," Mathematics of Operations Research 17, pp- 97-116, 1992.

[20] Hiriart-Urruty and C. Lemarechal, Convex Analysis and Minimization Algorithm II, Springer Verlag, New York, 1996.

[21] A. Iusem , " Augmented Lagrangian methods and proximal point methods for convex optimization", Minicurso I Workshop en Optimización, 8, Florianópolis, Brasil. Diciembre 1.997.

[22] R.T. Rockafellar, "Augmented lagrangians and Aplications of the proximal point algorithm in convex programing", Mathematics of Operations Researchs 1, pp. 97-116, 1.976.

[23] R.T. Rockafellar, "The multiplier method of Hestenes and Powel applied to convex programing", JOTA, vol 12, nro 6, pp. 555-562, 1.973.

[24] Tseng P. y Bertsekas D., "On the convergence of the exponencial multiplier method for convex programing", Mathematical programing, 60, pp 1-19, 1.993.

[25] Moreau J., "Proximité et dualité dans un espace Hilbertien", Bull. Soc. Math. France 93, pp. 97-116, 1.976.

[26] Martinet B., "Regularisatio Dínequations variationnelles per aproximations successives", Revue Française de Informatiqué et Recherche Operationelle 2, pp. 154-159, 1.970.

[27] R.T. Rockafellar, "Monotone operators and the proximal point algorithm in convex programing", SIAM J. on Control and Optimization 14, pp. 877-898, 1.976.

[28] Attouch H. e Wets R., "Ephigraphical analysis", Ann. Ins. Poincaré:, Anal. Non-lineaire, 1.989.

[29] Humes C. e Da Silva P., "Proximal point, separators and augmented lagrangians", Work paper, U.S.P. São Paolo, Brasil, 1.977.

[30] Constantin Udriste, "Convex Functions and Optimization Methods on Riemannian Manifolds," Kluweer Academis Publishers. Mathematics and its applications. Vol 297. 1994.

[31] O.P. Ferreira and P.R. Oliveira, "Proximal Point Algorithm on Riemannian Manifolds," IME. Brazil. 2000.

[32] E.A. Quiroz and P.R. Oliveira, "Proximal Point Method for a class of Bregman Distance on Hadamard Manifolds," 2005.

[33] R.A. Castillo, E. Hernández, and J.Campos " A Generalized Like-Distance in Convex Programming", International Mathematical Forum, 2, 2007, $N°$ 37, 1811-1830.