REPÚBLICA BOLIVARIANA DE VENEZUELA.
UNIVERSIDAD CENTROOCCIDENTAL LISANDRO ALVARADO.
MAESTRÍA EN CIENCIAS, MENCIÓN OPTIMIZACIÓN.

UN MÉTODO DE MULTIPLICADORES BASADOS EN SHIFTS CON UNA PENALIDAD NO COERCIVA-

Lic. José Valentín Santamaría.

Barquisimeto, 2004.

REPÚBLICA BOLIVARIANA DE VENEZUELA.
UNIVERSIDAD CENTROOCCIDENTAL LISANDRO ALVARADO.
MAESTRÍA EN CIENCIAS, MENCIÓN OPTIMIZACIÓN.

UN MÉTODO DE MULTIPLICADORES BASADOS EN SHIFTS CON UNA PENALIDAD NO COERCIVA-

Trabajo de Grado presentado como requisito parcial para optar al Título de Magister Scientiarum, Mención Optimización.

Autor: Lic. José V. Santamaría.
Tutor: Dr: Rómulo Castillo.

Barquisimeto, 2004.

DEDICATORIA

- A mi **Dios Todopoderoso**, porque El es el todo en mi vida.
- A mi amada y hermosa esposa **Liliana**, por ser la mujer de mi vida.
- A mis queridos hijos, **Jorge Luis** y **Liz Valentina**, porque ellos son la razón de todos mis logros.
- A mis admirados padres **Eloisa** y **Valentín**, porque que me lo dieron todo.

AGRADECIMIENTOS

La culminación de esta investigación fue posible por la ayuda y el apoyo de muchas personas al igual que la finalización de la Maestría en Ciencias, en virtud de lo cual quiero expresar mis palabras de agradecimiento:

- A **Rómulo Castillo**, mi tutor académico, por sugerirme el tema, orientarme en el trabajo y mostrar su sincera amistad.
- A los profesores: **Hugo Lara, Javier Hernández** y **Alirio Dávila**, por contribuir en mi formación académica.
- A mi esposa **Liliana**, por ser mi constante apoyo y por ser tan comprensiva.
- A mis hijos **Liz Valentina** y **Jorge Luis**, porque ellos me inspiran para triunfar.
- A mis padres **Eloisa** y **Valentín** porque me han apoyado incondicionalmente.
- A mi iglesia **Lirio de los Valles**, por sus constantes oraciones.
- A mis colegas **Adalys** y **Enedina**, por sus palabras de animo en todo tiempo.
- A **Inés**, por su apoyo.
- A todos mis amigos, especialmente **Geovani** y **Nelson**, por la sincera solidaridad.
- Por sobre todas las cosas a mi **Dios**, por ser mi Guiador, Consolador y Ayudador en todos los momentos de mi vida.

Resumen

En [9] Kiwiel presenta un método para resolver un problema de programación convexa, en tal sentido considera un método de punto proximal que usa una función de Bregman generalizada, el cual denomina BPM. De igual forma presenta un método de multiplicadores para resolver el problema dual, el cual denomina método de multiplicadores inexactos.

Un método similar presenta Castillo en [3], denominado método de multiplicadores basados en shifts, para el cual presenta dos algoritmos, uno usando una penalidad no coerciva de tipo AL1 y otro usando una penalidad coerciva de tipo AL2. En este trabajo implementamos este método de multiplicadores basados en shifts usando una penalidad θ considerada por Zang en [19], la cual es no coerciva y no la considera Kiwiel en su estudio, tampoco la considera Castillo, ya que θ no es estrictamente convexa y por tanto no es de tipo AL1, no obstante cumple todas las otras propiedades de AL1 y se puede aplicar el método.

Al aplicar este método con la penalidad θ, se puede observar que con algunas condiciones la sucesión $\{\mu^k\}$ converge, además la sucesión $\{x^k\}$ es acotada y converge a una solución óptima del problema. Este resultado es obtenido sin necesidad de usar dos condiciones que considera Castillo en la demostración de un teorema similar en [3].
Para probar la eficiencia computacional de este método que usa la penalidad no coerciva θ se realizó un programa computacional con el software Matlab, con el cual se resuelven algunos problemas tomados de [7] y [2], observándose buen desempeño, ya que obtiene buenas aproximaciones de las soluciones óptimas.

Palabras claves: penalidad, no coerciva, programación convexa, B-funcion, método de lagrangeano aumentado, método de punto proximal.

Índice General

INTRODUCCIÓN

Para resolver problemas de programación convexa, donde la función objetivo y las funciones asociadas a las restricciones son funciones convexas, se han usado los métodos de lagrangeano aumentado debido a su buen desempeño computacional y por su conexión con los métodos de punto proximal usados para resolver el problema dual.

Existen varias versiones de los métodos de lagrangeano aumentado, asi como las hay de los métodos de punto proximal. En [9] Kiwiel presenta un método de punto proximal denominado Bregman Proximal Minimization (BPM), el cual usa una función h de Bregman generalizada, es decir una generalización de una B-función usual, además esta función h puede ser no diferenciable, por lo cual en el BPM en vez de usar el gradiente de h se usan sus subgradientes. Asociado a este BPM Kiwiel presenta un método de multiplicadores para resolver el problema dual, el cual denomina método de multiplicadores inexactos, para ello usa una función de penalidad esencialmente suave. Esta función de penalidad es la monótona conjugada de h, además es una función coerciva cuyo dominio contiene a los números reales positivos. Es importante resaltar que Kiwiel aborda los problemas primal y dual de manera inversa a como se plantean usualmente, de allí que resuelva el primal con un método de punto proximal y el dual con un método de multiplicadores.

Castillo en [3] introduce métodos similares, para la resolución del problema primal presenta un método de multiplicadores denominado método de multiplicadores basados en shifts, en la aplicación de este método Castillo considera dos tipos de penalidad, a saber las AL1 que son no coercivas y las

AL2 que son coercivas. Una de las propiedades que se requieren de estas funciones de penalidad es que sean diferenciable. De igual manera presenta para el problema dual un método de punto proximal. Una diferencia notable entre los métodos de Kiwiel y Castillo es que Kiwiel no considera en su método de multiplicadores inexactos penalidades no coercivas, mientras Castillo si las considera en su método de multiplicadores basados en shifts.

El objetivo fundamental del presente trabajo es implementar el método de multiplicadores basados en shifts usado por Castillo en [3] con una penalidad específica θ, la cual es diferenciable y no coerciva. Para ello se hace un estudio de los métodos de lagrangeano aumentado y los de punto proximal y en particular se analizan de manera detallada los presentados por Kiwiel y Castillo en [9] y [3] respectivamente.

El método de multiplicadores basados en shifts con la penalidad no coerciva θ genera dos sucesiones, a saber $\{x^k\}$ y $\{\mu^k\}$. En este trabajo se establece bajo ciertas condiciones la convergencia de $\{\mu^k\}$, además se prueba un teorema similar a uno presentado por Castillo en [3] donde se establece la factibilidad ergódica, la complementaridad ergódica y la convergencia ergódica de una sucesión $\{\bar{x}^k\}$ de medias primal asociada a la sucesión $\{x^k\}$. En su teorema Castillo plantea estos resultados para las penalidades AL1 y AL2, pero demuestra sólo el caso AL2, dejando planteado el caso de AL1. La demostración que se realiza en esta investigación es con la penalidad θ, la cual cumple todas las propiedades de AL1, excepto la convexidad estricta.

En este mismo teorema se establece la acotación de la sucesión $\{x^k\}$, así como su convergencia a una solución óptima del problema primal sin necesidad de considerar dos condiciones adicionales que se imponen en el teorema presentado por Castillo en [3].

Con el fin de mostrar la eficiencia del método estudiado usando la penalidad no coerciva θ se realiza una implementación numérica para lo cual se hace un programa computacional, usando el software Matlab, para aplicarlo en la resolución de algunos problemas propuestos por Hock y Schittkowski en [7] y

Bazaraa y Shetty en [2]. Al aplicar el programa se observa buen desempeño, ya que se hallan las soluciones óptimas.

Este trabajo se ha dividido en cuatro capítulos. En el capítulo 1 se estudian los preliminares, los cuales estan constituidos por una serie de definiciones y proposiciones de análisis convexo que proveen el fundamento teórico de esta investigación. los temas abordados son funciones convexas, diferenciabilidad, subdiferenciabilidad y las propiedades de la conjugada.

En el capítulo 2 se hace un estudio de los métodos de lagrangeano aumentados y los de punto proximal. Se presentan algunas versiones del método de lagrangeano aumentado, a saber el método de multiplicadores clásico presentado por Rockafellar en [16], el método de lagrangeano aumentado considerado por Iusem en [8] y el método de multiplicadores basados en shifts considerados por Castillo en [3]. De igual modo se presenta el método de punto proximal asociado a la norma euclidiana y el asociado a una función de Bregman. En este capítulo se presenta un teorema interesante que permite relacionar un método de lagrangeano aumentado con un método de punto proximal.

En el capítulo 3 se estudia el método de punto proximal (BPM) presentado por Kiwiel en [9], el cual usa una función de Bregman generalizada, se hace un análisis detallado de este método y de la convergencia de la sucesión que genera, de igual forma se estudia el método de multiplicadores inexactos presentados por Kiwiel, así como el análisis de convergencia de las sucesiones que se generan de este método. También se presentan es este capítulo varias familias de penalidades y sus correspondientes conjugadas.

En el capítulo 4 se presenta la implementación numérica del método de multiplicadores basados en shifts con la penalidad no coerciva θ. En primer lugar se hace un estudio detallado de la función θ, se obtiene que es una función propia, convexa, cerrada, continua, esencialmente suave, no coerciva y cumple las propiedades de AL1, excepto la convexidad estricta. Con esta función se pueban varios resultados del método bajo estudio presentados por

Castillo en [3], entre estos resultados se prueba que la sucesión generada por el método de multiplicadores basados en shifts con la penalidad θ es acotada y converge a una solución del problema, esto se demuestra omitiendo dos condiciones impuesta por Castillo en un teorema similar. Finalmente en este capítulo se realiza una implementación numérica a través de la realización de un programa computacional y resolviendo algunos problemas.

Capítulo 1

PRELIMINARES

El propósito fundamental de este trabajo es hacer un estudio detallado acerca de la resolución de un problema de minimización de una función convexa con restricciónes asociadas a funciones convexas. Para resolver este tipo de problema se usan los métodos de multiplicadores o de lagrangeano aumentado, de los cuales existen varias versiones. Estos métodos usan una función de penalidad parametrizada. En esta investigación se usa el método de multiplicadores basados en shifts, descrito por Castillo en [3], usando una penalidad en particular la cual posee una diferencia sustancial con las penalidades que regularmente son usadas en los métodos de multiplicadores.

Dado que este problema corresponde a un problema de programación convexa, enunciaremos algunas definiciones y teoremas fundamentales sobre el análisis convexo, con el fin de facilitar la comprensión y el análisis de los temas que serán desarrollados en los próximos capítulos. El propósito es establecer una base teórica suficiente para describir los métodos de punto proximal y los de multiplicadores. En particular abordaremos el tema de funciones convexas, diferenciabilidad y subdiferenciabilidad, así como las propiedades de su conjugada. Cada uno de estos resultados pueden ser revisados en cualquier libro de análisis convexo, en particular se recomienda revisar el de Rockafellar (ver [16]) y el de Hiriart Urruty y Lemaréchal (ver [6]).

1-1 FUNCIONES CONVEXAS

En esta sección estudiaremos la noción de convexidad, la cual es una propiedad utilizada a lo largo del presente trabajo, de esta forma facilitaremos la compresión de los temas tratados.
Comenzaremos por definir a un conjunto convexo.

Definición 1.1.1 CONJUNTOS CONVEXOS
Un conjunto $A \subseteq \mathbb{R}^n$ *es convexo si y sólo si para cada* $x, y \in A$, $\alpha \in (0,1)$ *se cumple:*

$$\alpha x + (1-\alpha)y \in A.$$

Podemos decir que un conjunto es convexo si el segmento que une a cualquier par de puntos del conjunto está totalmente contenido en el conjunto.

Definición 1.1.2 EPIGRAFO DE UNA FUNCIÓN
Dada una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, *no identicamente igual a* $+\infty$, *el epigrafo de la función* f *es el conjunto no vacío*

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

Si consideramos el gráfico de la función, el cual es el conjunto:

$$\mathrm{graf}f = \{(x,r) \in \mathbb{R}^n \times \mathbb{R} : f(x) = r\},$$

luego es fácil verificar que el gráfico de la función f es un subconjunto del epigrafo de la misma.

A continuación presentamos la definición de una función convexa:

Definición 1.1.3 FUNCIÓN CONVEXA
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *es convexa si sólo si para cada* $x, y \in R^n$ *y* $\alpha \in (0,1)$ *se cumple:*

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

La función f es estrictamente convexa si la desigualdad anterior se verifica de manera estricta.
La función f es concáva si la función $-f$ es convexa.
Una función es convexa si dados dos puntos x, y de su dominio , el segmento que une a los puntos $(x, f(x))$ y $(y, f(y))$ está siempre por encima o coincide con el gráfico de la función. De igual forma es cóncava si el segmento en consideración está por debajo o coincide con el gráfico de la función.

La siguiente proposición relaciona la convexidad de una función con la convexidad del epigrafo de la mencionada función.

Proposición 1.1.1 *(ver [6], Proposición 1.1.6)*
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es convexa si sólo si epif es un conjunto convexo.

Definición 1.1.4 DOMINIO EFECTIVO
Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función, el dominio efectivo de f es el conjunto:

$$\mathcal{D}_f = \{x | \exists \mu \in \mathbb{R} : (x, \mu) \in epif\} = \{x | f(x) < +\infty\}.$$

El dominio efectivo de una función es el conjunto formado por todos los puntos donde la función toma un valor finito.

Definición 1.1.5 FUNCIÓN PROPIA
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es propia si el epif es no vacío y f en ningún momento toma el valor de $-\infty$, es decir existe un $x_0 \in \mathbb{R}^n$ tal que $f(x_0) < +\infty$ y $f(x) \neq -\infty \ \forall x \in \mathbb{R}^n$.

A continuación presentamos un concepto asociado a la continuidad de una función:

Definición 1.1.6 FUNCIÓN SEMICONTINUA INFERIOR
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *es semicontinua inferior si para cada* $x \in \mathbb{R}^n$ *se verifica*

$$\liminf_{y \to x} f(y) \geq f(x). \tag{1.1}$$

Notemos que una función continua en $\mathbb{R}^n$ es semicontinua inferior, ya que $\liminf_{y \to x} f(y) = f(x)$.

Definición 1.1.7 FUNCIÓN CERRADA
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *es cerrada si epi* f *es un conjunto cerrado en* $\mathbb{R}^n \times \mathbb{R}$.

La siguiente proposición relaciona las funciones semicontinuas inferiores con las funciones cerradas, resultando que la semicontinuidad inferior es una forma de caracterizar a las funciones cerradas.

Proposición 1.1.2 *(ver [16], Teorema 7.1)*
Sea una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, *las siguientes condiciones son equivalentes:*

a) f *es semicontinua inferior en* $\mathbb{R}^n$.

b) $\{x | f(x) \leq \alpha\}$ *es cerrado para cada* $\alpha \in \mathbb{R}$.

c) f *es cerrada.*

Definición 1.1.8 FUNCIÓN CLAUSURA
Sea f *una función propia y convexa, la clausura de* f *es la más grande función semi-continua inferior mayorizada por* f. *Es decir es la función cuyo epifrafo es la clausura del epigrafo de* f *en* $\mathbb{R}^{n+1}$.
La función clausura de f *es llamada la cápsula semicontinua inferior de* f *y se denota por* clf.

Sí f es una función cerrada se tiene que $\mathrm{cl} f = f$, ya que por la proposición 1.1.2 el conjunto $\{x | f(x) \leq \alpha\}$ es cerrado para cada $\alpha \in \mathbb{R}$, por lo cual epi f

es cerrado y entonces es igual a su clausura, por lo cual $\mathrm{cl} f = f$.
Sí f es una función impropia tal que $f(x) = -\infty$, para algún x, la clausura de f es definida como la función constante igual a $-\infty$.

El siguiente teorema generaliza el concepto de convexidad de una función, para el caso de n puntos de su dominio.

Teorema 1.1.3 DESIGUALDAD DE JENSEN *(ver [16] Teorema 4.3)*
Sea una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ entonces f es convexa si sólo si para cada $x_1, \ldots, x_n \in \mathbb{R}^n$ se verifica:

$$f(\lambda_1 x_1 + \ldots + \lambda_m x_m) \leq \lambda_1 f(x_1) + \ldots + \lambda_m f(x_m),$$

con $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0 \qquad y \qquad \lambda_1 + \ldots + \lambda_m = 1.$

Las siguientes dos proposiciones relacionan la convexidad de una función con el concepto de diferenciabilidad, a través de el uso del gradiente de la función.

Proposición 1.1.4 *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia tal que $\mathcal{D}_f$ es abierto y convexo, con f diferenciable en $\mathcal{D}_f$: Entonces f es convexa si sólo si $\forall x, y \in \mathcal{D}_f$ se verifica:*

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

donde $\nabla f(x)$ es el gradiente de f en x.

Proposición 1.1.5 *Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia tal que $\mathcal{D}_f$ es abierto y convexo, con f diferenciable en $\mathcal{D}_f$, entonces f es convexa si y sólo si $\forall x, y \in \mathcal{D}_f$ se tiene:*

$$\langle y - x, \nabla f(y) - \nabla f(x) \rangle \geq 0.$$

La condición de diferenciabilidad en una función a menudo es una propiedad fuerte y puede darse el caso que no necesariamente las funciones que consideremos sean diferenciables o posean gradientes. En este sentido consideramos un concepto relacionado con la diferenciabilidad, como lo es el concepto de subgradiente y subdiferencial.

Definición 1.1.9 SUBGRADIENTES
Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *una función propia y convexa, sea* $x \in \mathbb{R}^n$. *Un vector* $s \in \mathbb{R}^n$ *es un subgradiente de la función* f *en el punto* x *si:*

$$f(z) \geq f(x) + \langle z - x, s \rangle, \qquad \forall z \in \mathbb{R}^n. \tag{1.2}$$

Nótese que si f es diferenciable en x, por la proposición 1.1.4 el gradiente $\nabla f(x)$ es en particular un subgradiente de f en x.

Definición 1.1.10 SUBDIFERENCIAL
El conjunto de todos los subgradientes de f *en* x *es llamado subdiferencial de* f *en* x *y se denota por:*

$$\partial f(x) = \{s : f(z) \geq f(x) + \langle z - x, s \rangle, \quad \forall z \in \mathbb{R}^n\}.$$

Similarmente definimos para cada $\epsilon \geq 0$ *el* $\epsilon-$*subdiferencial como el conjunto:*

$$\partial_\epsilon f(x) = \{s : f(z) \geq f(x) + \langle z - x, s \rangle - \epsilon \quad \forall z \in \mathbb{R}^n\}.$$

Si $\epsilon = 0$ *se tiene que el* $\epsilon-$*subdiferencial coincide con el subdiferencial* $\partial_0 f = \partial f$.

Definición 1.1.11 FUNCIÓN ESENCIALMENTE ESTRICTAMENTE CONVEXA
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *es esencialmente estrictamente convexa si es estrictamente convexa en cada subconjunto convexo de* $\mathcal{D}_{\partial f}$.

Definición 1.1.12 FUNCIÓN INDICADORA
Sea C *un conjunto no vacío, la función indicadora del conjunto* C *es:*

$$i_C(x) = \begin{cases} 0 & si\ x \in C \\ \infty & de\ otra\ forma. \end{cases}$$

Definición 1.1.13 FUNCIÓN SOPORTE
Sea C *un conjunto no vacío, la función soporte del conjunto* C *es:*

$$i_C^*(\bar{x}) = \sup\{\langle \bar{x}, x\rangle : \quad x \in C\}.$$

Una definición equivalente a la diferenciabilidad para una función finita es presentada a continuación:

Definición 1.1.14 FUNCIÓN SUAVE
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es suave si y sólo si es finita y diferenciable en todo $\mathbb{R}^n$.

Introduciendo una condición adicional a una función suave, consideramos la definición siguiente:

Definición 1.1.15 FUNCIÓN ESENCIALMENTE SUAVE
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ propia y convexa es esencialmente suave si satisface las siguientes condiciones:

a) $\overset{\circ}{\mathcal{D}}_f \neq \emptyset$.

b) f *es diferenciable en* $\overset{\circ}{\mathcal{D}}_f$.

c) $\lim_{k\to\infty} |\nabla f(x_k)| = +\infty$ *con* $\{x_k\}$ *una sucesión en* $\overset{\circ}{\mathcal{D}}_f$ *que converge a un punto* x *de la frontera de* $\overset{\circ}{\mathcal{D}}_f$.

Si f es diferenciable en $\mathbb{R}^n$, se tiene que f es esencialmente suave, ya que $\overset{\circ}{\mathcal{D}}_f = \mathbb{R}^n$ y no existe una sucesión que converja a un punto de frontera y así la propiedad $\lim_{k\to\infty} |\nabla f(x_k)| = +\infty$ se cumple trivialmente.

Definición 1.1.16 DERIVADA DIRECCIONAL
Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, sea $x \in \mathcal{D}_f$. Definimos la derivada direccional de f en x con respecto al vector y de la siguiente manera

$$f^{'}(x; y) = \lim_{\lambda \downarrow 0} \frac{f(x + \lambda y) - f(x)}{\lambda}. \tag{1.3}$$

La siguiente proposición relaciona la derivada direccional de una función diferenciable en un punto de su dominio y el gradiente de esa función en el mencionado punto.

Proposición 1.1.6 *(ver [16], pag 213)*
Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, sea $x \in \mathcal{D}_f$, si f es diferenciable en x se verifica:

$$f^{'}(x; y) = \langle \nabla f(x), y \rangle, \quad \forall y. \tag{1.4}$$

La derivada direccional también puede relacionarse con un subgradiente de una función que no sea necesariamente diferenciable, tal como se muestra en la siguiente proposición.

Proposición 1.1.7 *(ver [16], Teorema 23.2)*
Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ una función propia, cerrada y convexa , sea $x \in \mathcal{D}_f$. Entonces $s \in \partial f(x)$ si sólo si:

$$f^{'}(x; y) \geq \langle s, y \rangle, \quad \forall y. \tag{1.5}$$

Además la clausura de $f^{'}(x; y)$ es una función convexa con respecto de y, es la función soporte del conjunto cerrado y convexo $\partial f(x)$.

Definición 1.1.17 FUNCIÓN DE RECESIÓN
Consideremos una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa, propia y cerrada en $x \in \mathcal{D}_f$. La función f'_∞ definida por:

$$d \in \mathbb{R}^n \mapsto f'_\infty(d) = \lim_{\lambda \to +\infty} \frac{f(x + \lambda d) - f(x)}{\lambda} \tag{1.6}$$

es llamada función de recesión de f.

Definición 1.1.18 DIRECCIÓN DE RECESIÓN
Una dirección $d \neq 0$ es una dirección de recesión de f si $f'_\infty(d) \leq 0$.

Definición 1.1.19 CONO DE RECESIÓN
Sea el conjunto $C \subset \mathbb{R}^n$, no vacío y convexo. Definimos el cono de recesión de C de la siguiente manera:

$$C^{'}_\infty = \{y \in \mathbb{R}^n : x + \lambda y \in C, con\ x \in C \quad y \quad \lambda \geq 0\}. \tag{1.7}$$

Dado que el propósito fundamental de este trabajo es presentar el método de multiplicadores basados en shifts con una penalidad no coerciva, debemos tomar en consideración el concepto de función coerciva.

Definición 1.1.20 FUNCIÓN COERCIVA
Una función $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ propia, cerrada y convexa es coerciva si:

$$\lim_{\|x\| \to \infty} f(x) = +\infty.$$

Consideremos

$$f'_\infty(d) = \lim_{\lambda \to +\infty} \frac{f(x + \lambda d) - f(x)}{\lambda}$$

si $0 \in \mathcal{D}_f$, hacemos $x = 0$, de donde obtenemos:

$$\begin{aligned} f'_\infty(d) &= \lim_{\lambda \to +\infty} \frac{f(\lambda d) - f(0)}{\lambda} \\ &= \lim_{\lambda \to +\infty} \frac{f(\lambda d)}{\lambda}. \end{aligned}$$

Tomamos $\lambda = \frac{1}{r}$ entonces se puede escribir:

$$f'_\infty(d) = \lim_{r \to 0} r f(d/r), \tag{1.8}$$

considerando $d = e = (1, 1, \ldots, 1)$ se tiene:

$$f'_\infty(e) = \lim_{x \to +\infty} \frac{f(x)}{x}.$$

Si $f'_\infty(e) = +\infty$ entonces f crece más rapidamente que una función lineal y denominamos a f 1-coerciva (o sencillamente coerciva).
Si $f'_\infty(e) < +\infty$ entonces f se aproxima a una función lineal en la dirección 1, denominamos a f 0-coerciva (o sencillamente no coerciva).

1-2 FUNCIÓN CONJUGADA

En esta sección se hace una descripción de la conjugada de una función convexa. El estudio de la conjugada es útil en el sentido que a través de ella se pueden establecer relaciones importantes entre el primal y el dual de un problema de optimización.

Definición 1.2.1 *Consideremos la función convexa $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. La conjugada de f es una función f^* dada por:*

$$s \in \mathbb{R}^n \mapsto f^*(s) = \sup_x \{\langle x, s\rangle - f(x) : x \in \mathcal{D}_f\}.$$

Observaciones

1. f^* es cerrada, propia y convexa sí y sólo sí f es propia.
2. La conjugada g^* de una función $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ cóncava y cerrada es definida como:

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

3. Si $f = -g$ entonces $g^* \neq -f^*$.

   En efecto:

$$\begin{aligned} f^*(s) &= \sup_x \{\langle x, s\rangle - f(x)\} \\ &= \sup_x \{\langle x, s\rangle + g(x)\} \\ &= \sup_x \{-(-\langle x, s\rangle - g(x))\} \\ &= -\inf_x \{-\langle x, s\rangle - g(x)\} \\ &= -g^*(-s). \end{aligned}$$

Proposición 1.2.1 *(Ver [18], Proposición 2.10)*
Consideremos 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 $(f \circ A)^* = f^* \circ (A^{-1})^*$.*
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* $\mathcal{D}_{f_1} \cap \mathcal{D}_{f_2} \neq \emptyset \quad y \quad \alpha \in (0,1)$ *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

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

Definición 1.2.2 FUNCIÓN COFINITA
Una función $f : \mathbb{R}^n \to \mathbb{R}$ *propia y convexa es cofinita si epi* f *no contiene semi-rectas no verticales.*

La siguiente proposición establece que la función de recesión de una función cofinita es igual a infinito para todo punto distinto del cero.

Proposición 1.2.2 *(ver [16], Teorema 13.3)*
Una función $f : \mathbb{R}^n \to \mathbb{R}$ *propia y convexa es cofinita si*

$$f_\infty'(y) = +\infty, \quad \forall y \neq 0.$$

Otra caracterización de las funciones cofinitas es que su conjugada es finita para todo $s \in \mathbb{R}^n$. Este resultado es establecido en la siguiente proposición.

Proposición 1.2.3 *(ver [16] Corolario 13.3.1)*
Una función $f : \mathbb{R}^n \to \mathbb{R}$ *propia y convexa es cofinita si*

$$f^*(s) = \sup_x \{\langle s, x\rangle - f(x)\} < \infty, \quad \forall s \in \mathbb{R}^n,$$

lo cual significa que una función f *es cofinita si sólo si* $\mathcal{D}_{f^*} = \mathbb{R}^n$.

Una relación importante entre una función convexa y su conjugada se establece a través de la siguiente proposición:

Proposición 1.2.4 DESIGUALDAD DE FENCHEL
Si $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *es una función propia y convexa; se verifica la siguiente desigualdad:*

$$\langle x, s\rangle \leq f(x) + f^*(s) \qquad \forall x, s \in \mathbb{R}^n.$$

Existe una interesante relación de dualidad entre una función convexa y su conjugada, a través de los subgradientes de las mencionadas funciones. De esta forma podemos decir que s es un subgradiente de f en x si sólo si x es un subgradiente de f^* en s. En tal sentido se muestra la siguiente proposición.

Proposición 1.2.5 *(ver [6], Proposición 6.1.2 ; [16], Teorema 23.5)*
Si $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *es una función convexa, propia y cerrada, entonces las siguientes afirmaciones son equivalentes:*

1. $s \in \partial f(x)$.
2. $f(x) + f^*(s) = \langle s, x\rangle$.
3. $x \in \arg\min\{f(\cdot) - \langle s, \cdot\rangle\}$.
4. $x \in \partial f^*(s)$.

Demostración . Probemos en primer lugar ($1 \Rightarrow 2$).
Sea $s \in \partial f(x)$, por lo cual:

$$\begin{aligned} f(y) &\geq f(x) + \langle s, y - x\rangle \qquad \forall y \in \mathbb{R}^n \\ &= f(x) + \langle s, y\rangle - \langle s, x\rangle \qquad \forall y \in \mathbb{R}^n, \end{aligned}$$

luego:

$$\langle s, x\rangle - f(x) \geq \langle s, y\rangle - f(y), \quad \forall y \in \mathbb{R}^n,$$

por lo tanto:

$$f^*(s) = \langle s, x\rangle - f(x),$$

es decir:

$$f^*(s) + f(x) = \langle s, x\rangle.$$

Probemos ahora $(2 \Rightarrow 3)$. Sean x, s, tal que:

$$f^*(s) + f(x) = \langle s, x \rangle,$$

luego:

$$f^*(s) = \langle s, x \rangle - f(x),$$

de donde:

$$\sup_{y \in \mathbb{R}^n} \{\langle s, y \rangle - f(y)\} = \langle s, x \rangle - f(x),$$

deduciéndose:

$$\begin{aligned} \langle s, x \rangle - f(x) &\geq \langle s, y \rangle - f(y) \qquad \forall y \in \mathbb{R}^n. \\ f(x) - \langle s, x \rangle &\leq f(y) - \langle s, y \rangle \qquad \forall y \in \mathbb{R}^n, \end{aligned}$$

por lo cual:

$$x \in \arg\min\{f(\cdot) - \langle s, \cdot \rangle\}.$$

Probemos ahora $(3 \Rightarrow 4)$.
Sea $x \in \arg\min\{f(\cdot) - \langle s, \cdot \rangle\}$, entonces

$$\begin{aligned} f(x) - \langle s, x \rangle &\leq f(y) - \langle s, y \rangle, \quad \forall y \in \mathbb{R}^n \\ \langle s, x \rangle - f(x) &\geq \langle s, y \rangle - f(y), \quad \forall y \in \mathbb{R}^n \\ f^*(s) &= \langle s, x \rangle - f(x). \end{aligned}$$

Sea $p \in \mathbb{R}^n$ arbitrario, luego

$$f^*(p) \geq \langle p, y \rangle - f(y), \quad \forall y \in \mathbb{R}^n,$$

en particular se verifica para $y = x$:

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

por lo cual: $x \in \partial f^*(s)$.

Finalmente probamos $(4 \Rightarrow 1)$.

Sí $x \in \partial f^*(s)$ se tiene:

$$\begin{aligned} f^*(p) &\geq f^*(s) + \langle x, p - s\rangle \qquad \forall p \in \mathbb{R}^n \\ &\geq f^*(s) + \langle x, p\rangle - \langle x, s\rangle \qquad \forall p \in \mathbb{R}^n, \end{aligned}$$

luego:

$$\begin{aligned} f^*(p) - \langle x, p\rangle &\geq f^*(s) - \langle x, s\rangle \qquad \forall p \in \mathbb{R}^n \\ \langle x, p\rangle - f^*(p) &\leq \langle x, s\rangle - f^*(s) \qquad \forall p \in \mathbb{R}^n, \end{aligned}$$

por lo cual:

$$(f^*)^*(x) = \langle x, s\rangle - f^*(s),$$

pero como f es cerrada se tiene:

$$\begin{aligned} f(x) &= \langle x, s\rangle - f^*(s) \\ f^*(s) &= \langle x, s\rangle - f(x), \end{aligned}$$

y así:

$$\begin{aligned} \langle x, s\rangle - f(x) &\geq \langle y, s\rangle - f(y) \qquad \forall y \in \mathbb{R}^n \\ f(y) - \langle y, s\rangle &\geq f(x) - \langle x, s\rangle \qquad \forall y \in \mathbb{R}^n \\ f(y) &\geq f(x) - \langle x, s\rangle + \langle y, s\rangle \qquad \forall y \in \mathbb{R}^n \\ f(y) &\geq f(x) + \langle s, y - x\rangle \qquad \forall y \in \mathbb{R}^n, \end{aligned}$$

de donde: $s \in \partial f(x)$. ∎

De la proposición anterior se pueden deducir los siguientes resultados:

$$\partial f^* = (\partial f)^{-1}. \tag{1.9}$$

$$\text{im } \partial f = \mathcal{D}_{\partial f^*}. \tag{1.10}$$

De esta manera queda establecida que el subdiferencial de la conjugada de una función convexa es igual a la inversa del subdiferencial de la función.

La proposición que se presenta a continuación establece que el gradiente de una función propia y convexa es continuo en su dominio y además la función es diferenciable si sólo si el conjunto subdiferencial está compuesto unicamente por el gradiente, es decir el conjunto subdiferencial es unitario.

Proposición 1.2.6 *(Ver [16], Teoremas 25.1 y 25.5)*
Si h es una función propia y convexa, entonces ∇h es continua en $\mathcal{D}_{\nabla h} \subset \mathring{\mathcal{D}}$, si $x \in \mathcal{D}_{\nabla h}$ entonces h es diferenciable en x si sólo si $\partial h(x) = \{\nabla h(x)\}$.

Demostración . Sea $e_1, e_2, \ldots, e_n$ las filas de la matriz identidad $(n \times n)$, usando ([16], Teorema 25.4) se tiene $h'(x, e_j) = \frac{\partial h}{\partial x_j}(x)$ es continuo, y por tanto ∇h es continuo en $\mathcal{D}_{\nabla h}$.
Supongamos que h es diferenciable en x , entonces por (1.4) tenemos:

$$h'(x, y) = \langle \nabla h(x), y \rangle.$$

Si g es un subgradiente de h en x, por (1.5), se tiene:

$$h'(x, y) = \langle \nabla h(x), y \rangle \geq \langle g, y \rangle \quad \forall y.$$

Si tomamos $-y$ tenemos

$$\langle \nabla h(x), -y \rangle \geq \langle g, -y \rangle,$$

entonces

$$\langle \nabla h(x), y \rangle \leq \langle g, y \rangle,$$

y así $\nabla h(x) = g$, por lo tanto $\partial h(x) = \{\nabla h(x)\}$.
Recíprocamente, supongamos que $\partial h(x) = \{\nabla h(x)\}$.
Definimos la función $g(y) = h(x + y) - h(x) - \langle \nabla h(x), y \rangle$.

Como h es convexa g es convexa, en efecto si $y_1, y_2 \in \mathcal{D}_g$ y $\alpha \in (0,1)$, tenemos que:

$$\begin{aligned}
g(\alpha y_1 + (1-\alpha)y_2) &= h(x + \alpha y_1 + (1-\alpha)y_2) - h(x) - \langle \nabla h(x), \alpha y_1 + (1-\alpha)y_2 \rangle \\
&= h(\alpha(x+y_1) + (1-\alpha)(x+y_2)) + \alpha(-h(x) - \langle \nabla h(x), y_1 \rangle) + \\
&\quad +(1-\alpha)(-h(x) - \langle \nabla h(x), y_2 \rangle) \\
&\leq \alpha g(y_1) + (1-\alpha) g(y_2).
\end{aligned}$$

Si $x^* \in \partial g(0)$ se tiene $g(y) \geq g(0) + \langle x^*, y - 0 \rangle = \langle x^*, y \rangle, \quad \forall y,$
luego

$$h(x+y) - h(x) - \langle \nabla h(x), y \rangle \geq \langle x^*, y \rangle$$

$$h(x+y) \geq h(x) + \langle \nabla h(x), y \rangle + \langle x^*, y \rangle$$

$$h(x+y) \geq h(x) + \langle \nabla h(x) + x^*, y \rangle.$$

Como y es arbitrario $(\nabla h(x) + x^*) \in \partial h(x)$, pero por hipótesis $\partial h(x) = \{\nabla h(x)\}$, luego $x^* = 0$ y $\partial g(0) = \{0\}$.
Debemos probar que $\partial g(0) = \{0\}$ implica $\lim_{y \to 0} \frac{g(y)}{|y|} = 0$.
Por la proposición 1.1.7 la clausura de $g'(0, \cdot)$ es el soporte de $\partial g(0)$, entonces $g'(0, \cdot) = 0$,
y así:

$$0 = g'(0, u) = \lim_{\lambda \downarrow 0} [g(\lambda u) - g(0)]/\lambda, \quad \forall u.$$

Como $g(0) = 0$, tenemos que $\lim_{\lambda \downarrow 0} g(\lambda u)/\lambda = 0$.
Definimos $f_\lambda(u) = g(\lambda u)/\lambda$, $\lambda > 0$ la cual es convexa, ya que g lo es, y decrece a 0 cuando $\lambda \to 0$.
Sea B la bola euclidiana, sea $\{a_1, a_2, \ldots, a_m\}$ una colección finita de puntos cuya cápsula convexa está incluida en B.
Cada $u \in B$ se puede expresar como combinación convexa

$$u = \lambda_1 a_1 + \ldots + \lambda_m a_m.$$

Usando la desigualdad de Jensen (Teorema 1.1.3) se tiene:

$$0 \leq f_\lambda(u) \leq \sum_{i=1}^{m} \lambda_i f_\lambda(a_i) \leq \max\{f_\lambda(a_i) : i = 1, \ldots, m\}.$$

Como f_λ decrece a 0 para cada i cuando $\lambda \downarrow 0$ se concluye que $f_\lambda(u)$ decrece a 0 uniformemente en $u \in B$ cuando $\lambda \downarrow 0$.
Dado $\epsilon > 0$ existe $\delta > 0$ tal que $g(\lambda u)/\lambda \leq \epsilon, \quad \forall \lambda \in (0, \delta], \quad \forall u \in B$.
Como cada vector y tal que $0 < |y| \leq \delta$ puede expresarse como $y = \lambda u$ con $\lambda = |y|$ y $u \in B$ tenemos $g(y)/|y| \leq \epsilon$ cuando $0 < |y| \leq \delta$.
Esto prueba que $\lim_{y \to 0} \frac{g(y)}{|y|} = 0$ y así h es diferenciable. ∎

En la proposición siguiente se establece que el conjunto subdiferencial de una función esencialmente suave es el gradiente para los puntos del interior del dominio y es vacío para los punto de la frontera, además se presenta una relación entre una función esencialmente estricta convexa y su conjugada, la cual es esencialmente suave.

Proposición 1.2.7 *(ver [16], Teoremas 26.1 y 26.3)*
Sea h una función cerrada, propia y convexa entonces:

i) h es esencialmente suave si sólo si:

$$\partial h(x) = \{\nabla h(x)\}, \quad \forall x \in \overset{\circ}{\mathcal{D}}$$

y para cada x en la frontera de $\mathcal{D}_h$ se tiene $\partial h(x) = \emptyset$.

ii) h es esencialmente estricta convexa si sólo si h^* es esencialmente suave.

Demostración .

i) Si h es esencialmente suave es diferenciable en $\overset{\circ}{\mathcal{D}}_h$, usando la proposición 1.2.6 se deduce que $\partial h(x) = \{\nabla h(x)\}, \ \forall x \in \overset{\circ}{\mathcal{D}}_h$.

Sea x en la frontera de $\mathcal{D}_h$, supongamos que $\partial h(x) \neq \emptyset$, entonces existe una sucesión $\{x^k\}$ que converge a x, tal que $\partial h(x)$ contiene al límite de la sucesión $\{\nabla h(x^k)\}$ y así h no es esencialmente suave, contradictorio con la hipótesis. Por lo cual $\partial h(x) = \emptyset$.

Recíprocamente, si $\partial h(x) = \{\nabla h(x)\}$, $\quad \forall x \in \overset{\circ}{\mathcal{D}}$ y $\partial h(x) = \emptyset$, $\forall x$ en la frontera de $\mathcal{D}_h$ se deduce que $\overset{\circ}{\mathcal{D}}_{\partial h} \neq \emptyset$, ya que h es propia. Usando la Proposición 1.2.6 se tiene que h es diferenciable en $\overset{\circ}{\mathcal{D}}_h$.

Supongamos que la propiedad (c) de la Definición 1.1.15 no es válida, entonces existe una sucesión $\{x^k\}$ convergente a x en la frontera de $\mathcal{D}_h$ tal que $\{\nabla h(x^k)\}$ es acotada, entonces existe una subsucesión $\{\nabla h(x^{k_l})\}$ convergente a un cierto x^*, entonces $x^* \in \partial h(x)$ y así $\partial h(x) \neq \emptyset$, contradictorio con la hipótesis, y (c) es válida. Por lo tanto h es esencialmente suave.

ii Por (1.9) se tiene $\partial h^* = (\partial h)^{-1}$, además por la parte (i) ∂h^* es univaluada si sólo si h^* es esencialmente suave. Entonces es suficiente probar que h es esencialmente estrictamente convexa si sólo si $\partial h(x_1) \cap \partial h(x_2) = \emptyset$, cuando $x_1 \neq x_2$.

Supongamos que h no es esencialmente estrictamente convexa, entonces existen $x_1, x_2,\ x_1 \neq x_2$ tal que para cierto $x = \lambda x_1 + (1 - \lambda) x_2, \quad 0 < \lambda < 1$ se tiene $\partial h(x) \neq \emptyset$ y $h(x) = \lambda h(x_1) + (1-\lambda)h(x_2)$.

Tomemos $s \in \partial h(x)$, y sea H el gráfico de la función $f(z) = h(x) + \langle s, z - x\rangle$, H es el hiperplano soporte de epi$h$ en el punto $(x, h(x))$.

Por otro lado $(x, h(x))$ es un punto del interior relativo de la linea que une a los puntos $(x_1, h(x_1))$ y $(x_2, h(x_2))$.

Como $s \in \partial h(x)$ se tiene $h(x_1) \geq h(x) + \langle s, x_1 - x\rangle = f(x_1)$.

Además

$$\begin{aligned}
f(x_1) &= h(x) + \langle s, x_1 - x\rangle \\
&= \lambda h(x_1) + (1-\lambda)h(x_2) + \langle s, x_1 - \lambda x_1 - (1-\lambda)x_2\rangle \\
&= \lambda h(x_1) + (1-\lambda)h(x_2) + (1-\lambda)\langle s, x_1 - x_2\rangle \\
&\geq \lambda h(x_1) + (1-\lambda)h(x) + (1-\lambda)\langle s, x_2 - x\rangle + (1-\lambda)\langle s, x_1 - x_2\rangle \\
&= \lambda h(x_1) + (1-\lambda)h(x) + (1-\lambda)\langle s, x_1 - x\rangle \\
&= \lambda h(x_1) + (1-\lambda)f(x_1).
\end{aligned}$$

Y así $f(x_1) \geq h(x_1)$ por lo cual $f(x_1) = h(x_1)$ y entonces $(x_1, h(x_1)) \in H$, similarmente se obtiene $(x_2, h(x_2)) \in H$, de donde $s \in \partial h(x_1)$ y $s \in \partial h(x_2)$, por lo cual $\partial h(x_1) \cap \partial h(x_2) \neq \emptyset$, lo cual es contradictorio, por lo tanto h es esencialmente estrictamente convexa.

Recíprocamente si h es esencialmente estrictamente convexa, supongamos que $s \in \partial h(x_1) \cap \partial h(x_2)$, con $x_1 \neq x_2$.
Sea $g(z) = \langle s, z \rangle - \mu$, donde $\mu = h^*(s)$. Consideremos a H, el gráfico de g, el cual es el hiperplano soporte no vertical de epi$h$.
H contiene a $(x_1, h(x_1))$, ya que si $s \in \partial h(x_1)$ por la Proposición 1.2.5 $h(x_1)+h^*(s) = \langle s, x_1 \rangle$ y así $h(x_1) = \langle s, x_1 \rangle - h^*(s) = \langle s, x_1 \rangle - \mu = g(x_1)$. De manera similar se prueba que $(x_2, h(x_2)) \in H$. Además la línea que une a ambos puntos también está contenida en H, por lo cual h no puede ser estrictamente convexa a lo largo del segmento que une a x_1 y x_2. Cada punto x en este segmento tiene un $s \in \partial h(x)$. De donde se deduce que h no es esencialmente estrictamente convexa, lo cual contradice la hipótesis y así $\partial h(x_1) \cap \partial h(x_2) = \emptyset$.

∎

Capítulo 2

MÉTODOS DE MULTIPLICADORES Y DE PUNTO PROXIMAL

En este capítulo estudiaremos los diversos métodos de lagrangeano aumentado, así como el método de punto proximal. Se establecerá una relación entre el método de multiplicadores que resuelve un problema primal y el método de punto proximal que resuelve el problema dual.

El estudio de estos métodos se justifican, ya que el propósito fundamental de este trabajo es estudiar un método específico de multiplicadores, a saber el método de multiplicadores basados en Shifts con penalidad no coerciva, considerado por Castillo en [3]. Un método similar presenta Kiwiel en [9], llamado método de multiplicadores inexactos es asociado con el método de minimización proximal con funciones de Bregman generalizadas.

El aporte que se pretende hacer con este trabajo es aplicar el método con una función de penalidad específica, la cual tiene como característica fundamental que es no coerciva y no es considerada por Kiwiel en su investigación.

La función que aplicaremos es usada por Zang en [19] en problemas de minimización, pero no como función de penalidad, sino como una aproximación a la función $q(t) = \max(0, t),\ t \in \mathbb{R}$.

2-1 MÉTODOS DE MULTIPLICADORES

Los métodos de multiplicadores transforman un problema de minimización en una secuencia de subproblemas, adicionando una función de penalidad que involucra a un $\mu \in R^m$ a la función objetivo de los subproblemas.

Una de las ventajas de los métodos de multiplicadores es su buen desempeño computacional. Además pueden relacionarse con los métodos de punto proximal que resuelven el problema dual asociado al problema primal.

Existen diversas versiones de estos métodos. En este trabajo se consideran cuatro tipos, a saber el método de multiplicadores clásicos presentados por Rockafellar en [16], el método de lagrangeano aumentado considerado por Iusem en [8], el método de multiplicadores basados en shifts considerados por Castillo en [3] y el método de multiplicadores inexactos, considerados por Kiwiel en [9].

Consideremos el problema primal de minimización:

$$(P) \qquad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & f_i(x) \leq 0 \\ & x \in X \subset \mathbb{R}^n. \end{array}$$

Donde f y f_i son funciones convexas y continuamente diferenciables en X para $i = 1, 2, \ldots, m$.
El método de lagrangeano aumentado consiste en resolver el problema de la forma:

$$x^{k+1} = argmin_x\{f(x) + \textstyle\sum_{i=1}^m \mu_i^k f_i(x)\},$$

donde $\mu_i^k \geq 0 \quad \forall i = 1, \ldots, m$.
También podemos escribir:

$$x^{k+1} = \arg\min_x \{L(x, \mu^k)\},$$

donde:

$$L(x,\mu^k) = \begin{cases} f(x) + \sum_{i=1}^m \mu_i^k f_i(x) & \text{si } \mu \geq 0 \\ +\infty & \text{de otra forma.} \end{cases} \tag{2.1}$$

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

$$\varphi(\mu) = \min_{x \in \mathbb{R}^n} L(x, \mu), \tag{2.2}$$

esta función φ es cóncava. Podemos considerar el problema dual asociado al problema (P).

$$(D) \qquad \begin{array}{ll} \text{maximizar} & \varphi(\mu) \\ \text{sujeto a} & \mu \geq 0. \end{array}$$

Para cada $x \in X, \bar{\mu} > 0$ se verifica $\varphi(\bar{\mu}) \leq L(x, \bar{\mu}) \leq \max_{\mu > 0} L(x, \mu)$, por lo tanto:

$$\min_{x \in X} L(x, \mu) \leq \max_{\mu > 0} L(x, \mu). \tag{2.3}$$

A partir de esta desigualdad podemos deducir un teorema importante de dualidad denominado teorema débil de dualidad:

Teorema 2.1.1 **Teorema débil de dualidad** *(ver [13], Lema 14.7)*

$$\max_{\mu > 0} \min_{x \in X} L(x, \mu) \leq \min_{x \in X} \max_{\mu > 0} L(x, \mu). \tag{2.4}$$

Demostración . Por (2.3) se tiene que:

$$\min_{x \in X} L(x, \mu) \leq \max_{\mu > 0} L(x, \mu).$$

Esta desigualdad es válida para cada $x \in X$ y $\mu > 0$; en particular vale si tomamos el μ que maximiza el lado izquierdo y el x que minimiza el lado derecho, por lo cual:

$$\max_{\mu>0} \min_{x\in X} L(x,\mu) \leq \min_{x\in X} \max_{\mu>0} L(x,\mu).$$

∎

Llamaremos *punto de silla* de la función L a un punto $(x^*, \mu^*) \in X \times \mathbb{R}_>$ que verifica la condición:

$$L(x^*,\mu) \leq L(x^*,\mu^*) \leq L(x,\mu^*) \ \forall x \in X, \mu > 0.$$

De esta condición se infiere que x^* es un minimizador de L cuando μ se fija en μ^*, de igual modo μ^* es un maximizador de L cuando x es fijado en x^*.

El siguiente teorema proporciona un resultado importante, al establecer una condición para que la solución óptima de un problema primal y la del dual asociado coincidan:

Teorema 2.1.2 **Teorema fuerte de dualidad** *(ver [13], Lema 14.8)* *La condición*

$$\max_{\mu>0} \min_{x\in X} L(x,\mu) = \min_{x\in X} \max_{\mu>0} L(x,\mu), \tag{2.5}$$

es válida si sólo si existe un par (x^*, μ^*) que satisfacen la condición de un punto de silla de L.

Demostración . Supongamos que (x^*, μ^*) es un punto de ensillatura de L, entonces:

$$L(x^*,\mu) \leq L(x^*,\mu^*) \leq L(x,\mu^*),$$

por lo cual:

$$\max_{\mu>0} L(x^*,\mu) \leq L(x^*,\mu^*) \leq \min_{x\in X} L(x,\mu^*),$$

además:

$$\min_{x\in X} \max_{\mu>0} L(x,\mu) \leq \max_{\mu>0} L(x^*,\mu).$$

Por otro lado:

$$\min_{x\in X} L(x,\mu^*) \leq \max_{\mu>0}\min_{x\in X} L(x,\mu).$$

Luego en virtud de las tres últimas desigualdades deducimos:

$$\min_{x\in X}\max_{\mu>0} L(x,\mu) \leq L(x^*,\mu^*) \leq \max_{\mu>0}\min_{x\in X} L(x,\mu).$$

La desigualdad en el otro sentido se debe al Teorema débil de dualidad, (Teorema 2.1.1) por lo que podemos concluir:

$$\min_{x\in X}\max_{\mu>0} L(x,\mu) = L(x^*,\mu^*) = \max_{\mu>0}\min_{x\in X} L(x,\mu).$$

Supongamos ahora que (2.5) es válida, entonces existe (x^*,μ^*) tal que:

$$\min_{x\in X}\max_{\mu>0} L(x,\mu) = L(x^*,\mu^*) = \max_{\mu>0}\min_{x\in X} L(x,\mu),$$

luego para cada $x \in X$ y $\mu > 0$, se verifica:

$$L(x^*,\mu) \leq \max_{\mu>0} L(x^*,\mu) = L(x^*,\mu^*) = \min_{x\in X} L(x,\mu^*) \leq L(x,\mu^*),$$

por tanto (x^*,μ^*) es un punto de silla de L. ∎

Este teorema garantiza que la función objetivo del problema primal evaluada en un mimizador de este problema coincide con la función objetivo del problema dual evaluada en un maximizador de este problema, este resultado es de gran utilidad para relacionar al método de lagrangeano aumentado que resuelven el primal y el método de punto proximal que resuelve el problema dual.

2-1-1 MÉTODO DE LAGRANGEANO AUMENTADO

Estudiemos ahora el método de Lagrangeano Aumentado, considerado por Iusem en [8].

Para $x \in \mathbb{R}^n$ definimos x^+ de la siguiente forma $x_j^+ = \max\{x_j, 0\}$. Tomamos un parámetro $c > 0$ y definimos el Lagrangeano aumentado $L_c : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ como:

$$L_c(x,\mu) = f(x) + \frac{1}{4c}\sum_{i=1}^{m}\{[(\mu_i + 2cf_i(x))^+]^2 - \mu_i^2\}.$$

Si calculamos el gradiente al Lagrangeano Aumentado, obtenemos:

$$\nabla L_c(x,y) = \nabla f(x) + \sum_{i=1}^{m}(\mu_i + 2cf_i(x))^+ \nabla f_i(x). \tag{2.6}$$

El método en consideración genera dos sucesiones $\{x^k\} \subset \mathbb{R}^n, \quad \{\mu^k\} \subset \mathbb{R}^m$ de la siguiente forma:
Dado $\mu^0 \in \mathbb{R}^m_+$ se obtiene:

$$x^k = \arg\min_{x\in\mathbb{R}^n} L_c(x,\mu^k). \tag{2.7}$$

$$\mu^{k+1} = (\mu_i^k + 2cf_i(x^k))^+. \tag{2.8}$$

Existen otras formas de usar el método de multiplicadores, por ejemplo si se introduce una función θ y un parámetro $r \in (0,1]$ y obtenemos:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto L_r(x,\mu) = f(x) + r\sum_{i=1}^{m}\mu_i\theta\left(\frac{f_i(x)}{r}\right). \tag{2.9}$$

Donde la función θ se le llama función de penalidad y al parámetro r se le llama parámetro de penalidad.

Existen diversas familias de penalidad con características muy diversas, por ejemplo las descritas por Bertsekas, por Ben-Tal y Zibulevsky, por Polyak-Teboulle y por Humes-Da Silva (ver [3]).
Definiremos la familia de penalidad no coerciva (AL1) y la familia de penalidad coerciva (AL2), consideradas por Castillo en ([3]).

FAMILIA DE PENALIDAD NO COERCIVA (AL1):

Consideremos, en primer lugar, una función $\theta : \mathbb{R} \to \mathbb{R}$ convexa y creciente tal que:

i) θ es estrictamente convexa en $\mathbb{R}$.

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

iii) $\theta'_{\infty}(1) = 1$ (es no coerciva ó 0-coerciva).

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

v) $\theta(\cdot)$ es acotada inferiormente.

vi) $\sup_{y>0}\{y - \theta(y)\}$ es finito.

Dado $\beta_i > 0$ y $\mu_i \in (0, \beta_i)$ para $i = 1, \dots, m$, definimos la función $\tilde{y}_i(\cdot)$ como:

$$\mu_i \in (0, \beta_i) \mapsto \tilde{y}_i(\mu_i) = (\theta')^{-1}\left(\frac{\mu_i}{\beta_i}\right), \quad \text{para } i = 1, \dots, m. \tag{2.10}$$

Dados $\beta \in \mathbb{R}^m_>$, $r \in (0, 1]$ fijos, definimos una penalidad $P_{\beta,r}$ perteneciente a la familia de penalidad AL1, como la función $P_{\beta,r} : \mathbb{R}^m \times (0, \beta) \to \mathbb{R}$ definida como sigue:

$$\begin{aligned} P_{\beta,r}(y, \mu) &= \sum_{i=1}^{m} P_{\beta_i,r}(y_i, \mu_i) \\ &= \sum_{i=1}^{m} \beta_i r \left[\theta\left(\frac{y_i}{r} + \tilde{y}_i(\mu_i)\right) - \theta(\tilde{y}_i(\mu_i))\right], \end{aligned} \tag{2.11}$$

donde $\tilde{y} = \tilde{y}(\mu)$ satisface $\theta'(\tilde{y}_i) = \frac{\mu_i}{\beta_i} < 1$ para $i = 1, \dots, m$.

En el capítulo 4, se presentará una función de penalidad θ, que cumple con las condiciones de la familia AL1, excepto la convexidad estricta. Con esta función de penalidad se aplicará el método de multiplicadores basados en shifts con penalidad no coerciva. Una de las ventajas de esta función que se presentará es que se puede tener la acotación de la sucesión $\{x^k\}$, generada por el mencionado método, además cada punto límite es una solución del problema primal, estos resultados serán obtenidos sin necesidad de imponer dos condiciones que usa Castillo en [3] para obtener el mismo resultado.

FAMILIA DE PENALIDAD COERCIVA (AL2):

Consideremos, en primer lugar, una función $\theta : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ convexa, propia, cerrada, creciente y diferenciable con $\mathrm{ri}\mathcal{D}_\theta = (-\infty, b)$, donde $b \in \mathbb{R}_>$ tal que:

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

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

iii) $\theta'_\infty(1) = \infty$ (es coerciva ó 1-coerciva).

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

v) $\theta(\cdot)$ es acotada inferiormente.

Dado $\beta_i \leq 1$ y $\mu_i \in \mathbb{R}_>$, para $i = 1, \ldots, m$ definimos la función $\tilde{y}_i(\cdot)$, como sigue:

$$\mu_i \in \mathbb{R}_> \mapsto \tilde{y}_i(\mu_i) = (\theta')^{-1}(\mu_i). \tag{2.12}$$

De esta forma podemos escribir $\theta'(\tilde{y}_i) = \mu_i$.
La función de penalidad $P_{\beta,r}$ perteneciente a la familia de penalidad AL2 se define de igual modo que en (2.11).

2-1-2 MÉTODO DE MULTIPLICADORES BASADOS EN SHIFTS

Con las familias de penalidad definidas anteriormente Castillo en [3] presenta un método de multiplicadores basados en shifts.
Un shifts consiste en trasladar el punto $(\tilde{y}, \theta(\tilde{y}))$ al origen, además $\theta'(\tilde{y}) = \mu$, por lo cual al introducir el shifts se obtiene que la penalidad $P(\tilde{y}, \mu)$ pase por el origen y tenga derivada μ.

MÉTODO DE MULTIPLICADORES BASADOS EN SHIFTS PARA PENALIDAD NO COERCIVA (AL1):

Este es el método que aplicaremos, pero con una penalidad específica, que será considerada en el capítulo 4.

Dados $\beta \geq 1,\ r \in (0,1]$, consideremos la función:

$$x \in \mathbb{R}^n, \mu \in (0,\beta) \mapsto L_{\beta,r}(x,\mu) = f(x) + \sum_{i=1}^{m} \beta_i r \left[\theta\left(\frac{f_i(x)}{r} + \tilde{y}_i(\mu_i)\right) - \theta(\tilde{y}_i(\mu_i)) \right], \tag{2.13}$$

donde θ es una penalidad de tipo AL1 y $\tilde{y}$ es tal que satisface $\theta^{'}(\tilde{y}_i(\mu_i)) = \frac{\mu_i}{\beta_i} < 1$ para $i = 1, \ldots, m$.
Usando esta función podemos generar el siguiente algoritmo:

Algoritmo 2.1.1 *Dados:* $\beta^0 \geq e,\ r^0 = 1,\ \mu^0 \in (0, \frac{\beta^0}{2})$ *y* $\tilde{y}^0$ *tal que* $\theta^{'}(\tilde{y}^0) = \frac{\mu^0}{\beta^0}$ *para* $k = 0, 1, \ldots$
Hallar:

$$x^{k+1} \in \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^{m} \theta_i^k \left(f_i(x)\right) \right\}, \tag{2.14}$$

donde $y_i = f_i(x) \in \mathbb{R}$ *y* $\theta_i^k(y_i) = \beta_i^k r^k \left[\theta\left(\frac{y_i}{r^k} + \tilde{y}_i^k\right) - \theta(\tilde{y}_i^k)\right].$
Actualizar:

$$\mu_i^{k+1} = \beta_i^k \theta^{'} \left(\frac{f_i(x^{k+1})}{r^k} + \tilde{y}_i^k \right), \quad \textit{para } i = 1, \ldots, m. \tag{2.15}$$

Para $i = 1, \ldots, m$; *sí* $\mu_i^{k+1} > \frac{\beta_i^k}{2}$ *entonces* $\beta_i^{k+1} = 2\beta_i^k$.
Calcular $\tilde{y}_i^{k+1}$ *tal que* $\theta^{'}(\tilde{y}_i^{k+1}) = \frac{\mu_i^{k+1}}{\beta_i^{k+1}}$,
sí $\mu_i^{k+1} \leq \frac{\beta_i^k}{2}$ *entonces* $\beta_i^{k+1} = \beta_i^k$ *y* $\tilde{y}_i^{k+1} = \tilde{y}_i^k + \frac{y_i^{k+1}}{r^k}$.
Escoger: $r^{k+1} \leq r^k$.

Con respecto a este algoritmo podemos hacer las siguientes observaciones:

- Por la definición de $\tilde{y}_i^{k+1}$, se tiene $\theta^{'}(\tilde{y}_i^{k+1}) = \frac{\mu^{k+1}}{\beta_i^{k+1}}$, por la actualización del multiplicador (2.15), tenemos: $\frac{\mu_i^{k+1}}{\beta_i^k} = \theta^{'}\left(\frac{y_i^{k+1}}{r^k} + \tilde{y}_i^k\right)$, por lo cual si $\beta_i^{k+1} = \beta_i^k$ se deduce que $\tilde{y}_i^{k+1} = \tilde{y}^k + \frac{y^{k+1}}{r^k}$.
- Por la definición de x^{k+1}, por las propiedades del subdiferencial, por la definición de L y por (2.15), tenemos:

$$\partial_x L_{\beta,r}(x^{k+1}, \mu^k) \;\; = \;\; \partial f(x^{k+1}) + \sum_{i=1}^{m} \beta_i^k \theta^{'} \left(\frac{f_i(x^{k+1})}{r} + \tilde{y}_i \right) \partial f_i(x^{k+1})$$

$$
\begin{aligned}
&= \partial f(x^{k+1}) + \sum_{i=1}^{m} \mu_i^{k+1} \partial f_i(x^{k+1}) \\
&= \partial_x L(x^{k+1}, \mu^{k+1}),
\end{aligned}
$$

por lo tanto:

$$
0 \in \partial_x L_{\beta,r}(x^{k+1}, \mu^k) \;\Leftrightarrow\; 0 \in \partial_x L(x^{k+1}, \mu^{k+1}). \tag{2.16}
$$

- Como θ es estrictamente creciente, tenemos que $\theta'(x) > 0,\ \forall x$, por otro lado $\mu^0 > 0$, por lo cual $\mu^k > 0$ para todo k.
- Como la sucesión $\{\mu^k\}$ es acotada, el número de veces que aumenta β^k es finito.
- Si $f_i(x^k) = 0$, para algún i y algún k, entonces $\mu_i^{k+1} = \beta_i^k \theta'(\tilde{y}_i^k) = \mu_i^k$.

Este algoritmo es el que será usado en nuestro análisis, ya que usa una penalidad no coerciva.

MÉTODO DE MULTIPLICADORES BASADOS EN SHIFTS PARA PENALIDAD COERCIVA (AL2):

En este método se omite el parámetro β.
Para $r \in (0, 1]$ consideremos la función penalizada:

$$
x \in \mathbb{R}^n, \mu \in \mathbb{R}^m_> \mapsto L_r(x, \mu) = f(x) + \sum_{i=1}^{m} r \left[\theta\left(\frac{f_i(x)}{r} + \tilde{y}_i(\mu_i) \right) - \theta(\tilde{y}_i(\mu_i)) \right],
$$

donde θ pertenece a la familia AL2 y $\tilde{y}$ satisface $\theta'(\tilde{y}_i(\mu_i)) = \mu_i$, para $i = 1, \ldots, m$.
Definimos el siguiente algoritmo:

Algoritmo 2.1.2 *Dados* $r^0 = 1,\ \mu^0 = \theta'(0),\ \tilde{y}^0 = 0$.
Para $k = 0, 1, \ldots$
Hallar:

$$
x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \{ f(x) + \sum_{i=1}^{m} \theta_i^k(f_i(x)) \}, \tag{2.17}
$$

donde $y_i \in \mathbb{R} \mapsto \theta_i^k(y_i) = r^k \left(\theta\left(\frac{y_i}{r^k} + \tilde{y}_i\right) - \theta(\tilde{y}_i^k)\right)$ *para todo* $i = 1, \ldots, m$.
Calcular:

$$\tilde{y}^{k+1} = \tilde{y}^k + \frac{y^{k+1}}{r^k}, \tag{2.18}$$

donde $y_i^{k+1} = f_i(x^{k+1})$.
Actualizar: $\mu_i^{k+1} = \theta'(\tilde{y}_i^{k+1})$ *para* $i = 1, \ldots, m$.
Escoger: $r^{k+1} \leq r^k$.

La diferencia entre este algoritmo y el AL1 es que las penalidades que usa AL1 son no coercivas y necesitan incrementar el parametro β, mientras las penalidades de AL2 son coercivas y no usan el parámetro β.
Las observaciones realizadas al algoritmo anterior son válidas también para este.

2-2 MÉTODO DE PUNTO PROXIMAL

Asociado al método de lagrangeano aumentado surge el método de punto proximal, tal método resuelve el problema dual (D).

Existen varias versiones de este método, la versión más sencilla es considerada por Iusem en [8], la cual suma a la función objetivo una función asociada a la norma euclidiana.

Otra versión es estudiada por Iusem en [8] y por Kiwiel en [9], la cual usa la Distancia asociada a una función de Bregman, que serán definidas en el presente capítulo.

Una versión más general es presentada por Kiwiel en [9], para ello define una generalización de la funciones de Bregman y las usa en el método que denomina Minimización Proximal de Bregman (BPM). Un estudio detallado de este método se presentará en el siguiente capítulo.

El método de punto proximal en su versión más sencilla resuelve un problema del siguiente tipo:

$$(D1) \qquad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & x \in \bar{X} \subset \mathbb{R}^n \end{array}$$

Donde f es una función convexa y continua en $\bar{X}$, además $X \subset \mathbb{R}^n$ es abierto y convexo, $\bar{X}$ es la clausura de X .

Aplicar el método consiste en generar una sucesión $\{x^k\} \subset \mathbb{R}^n$ de la siguiente manera:

$$x^0 \in X \subset \mathbb{R}^n. \tag{2.19}$$

$$x^{k+1} = \arg\min_{x \in \bar{X}} \{f(x) + \lambda_k \|x - x^k\|^2\}, \tag{2.20}$$

donde λ_k es un número real que satisface $0 < \lambda_k < \tilde{\lambda}$ para algún $\tilde{\lambda} > 0$.
Esta sucesión está bien definida, en efecto:
Procedemos por inducción:
Sea $f_k(x) = f(x) + \lambda_k \|x - x^k\|^2$.
Como

$$\lim_{\|x\| \to \infty} \lambda_k \|x - x^k\|^2 = \infty,$$

entonces:

$$\lim_{\|x\| \to \infty} f_k(x) = \infty$$

Por lo cual la minimización en (2.20) se reduce a un conjunto compacto, ya que f_k es continua, así f_k tiene mínimo. Como f es convexa y $\lambda_k \|x - x^k\|^2$ es estrictamente convexa, se tiene que f_k es estrictamente convexa. Por lo tanto

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

es único, y así la sucesión $\{x^k\}$ está bien definida.

Para establecer la convergencia de este método consideremos en primer lugar la siguiente definición, presentada por Iusem.

Definición 2.2.1 *(ver [8], Sección 2)*
Una sucesión $\{x^k\} \subset \mathbb{R}^n$ es llamada Fejér convergente a un conjunto $U \subset \mathbb{R}^n$ con respecto a la norma Euclidiana si.

$$\|x^{k+1} - u\| \leq \|x^k - u\|, \quad \forall k \geq 0, \ \forall u \in U. \tag{2.21}$$

La siguiente proposición relaciona la convergencia de Fejer y la acotación de una sucesión.

Proposición 2.2.1 *(ver [8], Proposición 2.1)*
Si la sucesión $\{x^k\} \subset \mathbb{R}^n$ es Fejér convergente a $U \neq \emptyset$, entonces $\{x^k\}$ es acotada. Si un punto de acumulación x de $\{x^k\}$ pertenece a U, entonces

$$x = \lim_{k\to\infty} x^k.$$

Demostración . (2.21) implica que para algún $u \in U$ se verifica:

$$\|x^k - u\| \leq \|x^0 - u\|, \quad \forall k \geq 0,$$

lo cual quiere decir que la sucesión $\{x^k\}$ está contenida en la bola de centro u y radio $\|x^0 - u\|$, por lo tanto es acotada.
Sea $x \in U$ un punto de acumulación de $\{x^k\}$, como $\{x^k\}$ es acotada poseé una subsucesión $\{x^{jk}\}$ convergente, tal que:

$$\lim_{k\to\infty} x^{jk} = x.$$

Por (2.21) se tiene que la sucesión $\{\|x^k - x\|\}$ es decreciente y no negativa, además posee una subsucesión $\{\|x^{jk} - x\|\}$, la cual converge a 0.

Por lo tanto $\{\|x^k - x\|\}$ converge a 0, y así:

$$\lim_{k\to\infty} \|x^k - x\| = 0,$$

lo que implica que:

$$\lim_{k\to\infty} x^k = x.$$

■

Una característica importante de este método de punto proximal que estamos estudiando es que podemos obtener la convergencia de la sucesión $\{x^k\}$, cuando la función objetivo es convexa y continuamente diferenciable, además la sucesión converge a un minimizador del problema, este resultado es establecido en la siguiente proposición:

Proposición 2.2.2 *(Ver [8], Teorema 2.1)*
Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función convexa y continuamente diferenciable. Supongamos que el conjunto U de minimizadores de f es no vacío, entonces la sucesión $\{x^k\}$ generada por (2.20) converge a un punto $x^* \in U$.

Demostración . En primer lugar tenemos que la sucesión está bien definida, probemos que:

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2, \quad \forall k \geq 0, \quad \forall \bar{x} \in U.$$

En efecto:
Sea $\bar{x} \in U$ arbitrario.

$$\begin{aligned} \|x^k - \bar{x}\|^2 &= \|x^k - x^{k+1} + x^{k+1} - \bar{x}\|^2 \\ &= \|x^k - x^{k+1}\|^2 + \|x^{k+1} - \bar{x}\|^2 + \\ &\quad + 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x}\rangle. \end{aligned} \tag{2.22}$$

Como x^{k+1} resuelve (2.20) se deduce:

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

por lo cual:

$$\nabla f(x^{k+1}) = 2\lambda_k(x^k - x^{k+1}). \tag{2.23}$$

Como f es convexa usamos la Proposición 1.1.4, consideramos además (2.22) y (2.23) para deducir:

$$\begin{aligned} \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 - \|x^{k+1} - \bar{x}\|^2 &= 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x}\rangle \\ &= \frac{1}{\lambda_k}\langle \nabla f(x^{k+1}), x^{k+1} - \bar{x}\rangle \\ &\geq \frac{1}{\lambda_k}[f(x^{k+1}) - f(\bar{x})]. \end{aligned} \tag{2.24}$$

Por otro lado $f(\bar{x}) \leq f(x^{k+1})$ porque $\bar{x}$ es un minimizador de f, por lo tanto de (2.24) se deduce:

$$0 \leq \|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2, \tag{2.25}$$

de lo cual resulta que la sucesión $\{x^k\}$ es Fejér convergente a U, ya que:

$$0 \leq \|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2. \tag{2.26}$$

Consideremos la sucesión $\{\|x^k - \bar{x}\|\}$, la cual es decreciente y no negativa, por lo tanto es convergente. Usando (2.25) se obtiene:

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0,$$

y así:

$$\lim_{k\to\infty} (x^{k+1} - x^k) = 0. \tag{2.27}$$

Sea x^* un punto de acumulación de $\{x^k\}$ y sea $\{x^{jk}\}$ una subsucesión de $\{x^k\}$ que converge a x^*.Usando (2.23) se obtiene:

$$\nabla f(x^{jk+1}) = 2\lambda_{jk}(x^{jk} - x^{jk+1}). \tag{2.28}$$

Por (2.27) se tiene que:

$$\lim_{k\to\infty} x^{jk+1} = \lim_{k\to\infty} x^{jk} = x^*.$$

Tomando límite en (2.28) cuando $k \to \infty$, usando $\lambda_k \leq \tilde{\lambda}$ y considerando que f es continuamente diferenciable obtenemos que:

$$\nabla f(x^*) = 0,$$

como f es convexa se tiene que $x^* \in U$.
Considerando que $\{x^k\}$ es Fejér convergente, por la Proposición 2.2.1 se deduce que:

$$\lim_{k\to\infty} x^k = x^*.$$

■

El siguiente teorema muestra una relación de dualidad entre el método de lagrangeano aumentado para el problema primal y el método de punto proximal para el problema dual.

Teorema 2.2.3 *(Ver [8], Teorema 7.1)*
Sea la sucesión $\{\bar{y}^k\} \subset \mathbb{R}^m$ generada por el método de punto proximal para

$\min_{y\in\mathbb{R}^m}\{-\varphi(y)\}$, *donde la función* φ *es definida por (2.2) y además* $\lambda_k = \frac{1}{4c}$ *y sea* $\{y^k\} \subset \mathbb{R}^m$ *la sucesión generada por (2.8). Si* $\bar{y}^0 = y^0$ *se tiene que* $\bar{y}^k = y^k, \quad \forall k \geq 0$.

Demostración . Procedemos por inducción.
Para $k = 0$, ia igualdad se verifica por hipótesis.
Supongamos que se cumple: $\bar{y}^k = y^k$ y probemos que $\bar{y}^{k+1} = y^{k+1}$.
Como

$$\bar{y}^{k+1} = \arg\min_{y\in\mathbb{R}^m}\{-\varphi(y) + \frac{1}{4c}\|y - y^k\|^2\},$$

entonces:

$$\frac{1}{2c}(\bar{y}^{k+1} - y^k) \in \partial\varphi(\bar{y}^{k+1}), \tag{2.29}$$

luego:

$$\frac{1}{2c}\langle\bar{y}^{k+1} - y^k, \bar{y}^{k+1} - y\rangle \leq \varphi(\bar{y}^{k+1}) - \varphi(y), \quad \forall y \geq 0,$$

por lo cual:

$$\varphi(y) \leq \varphi(\bar{y}^{k+1}) - \frac{1}{2c}\langle\bar{y}^{k+1} - y^k, \bar{y}^{k+1} - y\rangle, \quad \forall y \geq 0. \tag{2.30}$$

Como (2.30) es determinado unicamente por $\bar{y}^{k+1}$ es suficiente probar que (2.30) es verificado sustituyendo $\bar{y}^{k+1}$ por y^{k+1} Si usamos (2.23),(2.6),(2.8) y (2.1) se obtiene:

$$\begin{aligned} 0 = \nabla_x L_c(x^k, y^k) &= \nabla f(x^k) + \sum_{i=1}^{m}(y_i^k + 2cf_i(x^k))^+\nabla f_i(x^k) \\ &= \nabla f(x^k) + \sum_{i=1}^{m} y_i^{k+1}\nabla f(x^k) \in \partial_x L(x^k, y^{k+1}), \end{aligned} \tag{2.31}$$

implicando que x^k es un minimizador de $L(\cdot, y^{k+1})$ y por (2.2) se obtiene:

$$\varphi(y^{k+1}) = f(x^k) + \sum_{i=1}^{m} y_i^{k+1} f_i(x^k). \tag{2.32}$$

Por otro lado:

$$y_i^{k+1} - y_i^k = \max\{-y_i^k, 2cf_i(x^k)\},$$

lo cual implica:

$$y_i^{k+1}(y_i^{k+1} - y_i^k) = 2cy_i^{k+1} f_i(x^k). \tag{2.33}$$

Para cada $u \geq 0$ tenemos:

$$\begin{aligned}\frac{1}{2c}(y_i^{k+1}-y_i^k)(y_i^{k+1}-u_i) &= y_i^{k+1}f_i(x^k)-\frac{1}{2c}(y_i^{k+1}-y_i^k)u_i\\ &\leq y_i^{k+1}f_i(x^k)-u_if_i(x^k). \qquad (2.34)\end{aligned}$$

Por lo cual de (2.32) y (2.34) resulta:

$$\begin{aligned}\varphi(y^{k+1})-\frac{1}{2c}\langle y^{k+1}-y^k,y^{k+1}-u\rangle &\geq f(x^k)+\sum_{i=1}^m y_i^{k+1}f_i(x^k)-\\ &\quad -\sum_{i=1}^m y_i^{k+1}f_i(x^k)+\sum_{i=1}^m u_if_i(x^k)\\ &= L(x^k,u)\\ &\geq \min_{x\in\mathbb{R}^n} L(x,u)=\varphi(u). \qquad (2.35)\end{aligned}$$

Por lo tanto (2.30) es verificado para y^{k+1}, de donde obtenemos $y^{k+1}=\bar{y}^{k+1}$, y así la inducción es completada obteniéndose $y^k=\bar{y}^k,\ \forall k\geq 0$. ∎

En el método de punto proximal podemos introducir una versión usando funciones de Bregman, consideradas por Iusem en [8], las cuales se definen como sigue:

Definición 2.2.2 FUNCIONES DE BREGMAN *(ver [8], Sección 9)*
Sea $S\subset\mathbb{R}^n$ un conjunto abierto y convexo y $\bar{S}$ su clausura. Consideremos una función h definida en $\bar{S}$ y sea $D_h:\bar{S}\times S\to\mathbb{R}$ definida por:

$$D_h(x,y)=h(x)-h(y)-\langle\nabla h(y),x-y\rangle, \qquad (2.36)$$

h es llamada función de Bregman, (D_h es la Bregman distancia inducida por h) si se verifica lo siguiente:

i) h es estrictamente convexa en S.

ii) h es continua en $\bar{S}$.

iii) Para cada $\alpha\in\mathbb{R}$ los conjuntos parciales de nivel

$$\mathcal{L}_1(y,\alpha)=\{x\in\bar{S}:D_h(x,y)\leq\alpha\} \quad y \quad \mathcal{L}_2(x,\alpha)=\{y\in S:D_h(x,y)\leq\alpha\},$$

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

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

v) Si $\{x^k\} \subset \bar{S}$ *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* $\lim_{k\to\infty} x^k = y^*$.

A la función h también se le llama B-función, y al conjunto S se le denomina la zona de h.

Existen dos subclases de funciones de Bregman que requieren de condiciones adicionales a las mencionadas anteriormente. Las cuales se denominan coerciva en la frontera y coerciva en la zona.

Decimos 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

$$\lim_{k\to\infty} \nabla h(y^k)^t (x - y^k) = -\infty, \quad \forall x \in S.$$

Una función h de Bregman es coerciva en la zona si para cada $y \in \mathbb{R}^n$, existe $x \in S$, tal que $\nabla h(x) = y$.

Proposición 2.2.4 *Si* h *es una función de Bregman con zona* S*, entonces*

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

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

c) $D_h(\cdot, y)$ *es estrictamente convexa para cada* $y \in S$.

Demostración .

a) Usando (2.36) tenemos

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

b)

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

c) Como h es estrictamente convexa D_h también lo es.

∎

El método de punto proximal asociado a una función de Bregman se obtiene sustituyendo la función $\|\cdot\|$ por la distancia D_h asociada a una función h de Bregman, a través de la siguiente sucesión:

$$x^0 \in \bar{X} \subset \mathbb{R}^n. \tag{2.37}$$

$$x^{k+1} = \arg\min_{x\in\bar{X}}\{f(x) + D_h(x,x^k)/c_k\}, \tag{2.38}$$

donde D_h es la distancia asociada a una función h de Bregman, además $\{c_k\}$ es una sucesión de números positivos.

El siguiente teorema establece la convergencia de la sucesión generada por (2.38).

Teorema 2.2.5 *(Ver [8], Teorema 10.1)*
Si el problema $D1$ tiene solución y h es una función de Bregman coerciva en la frontera, entonces la sucesión generada por (2.37), (2.38) converge a una solución x^* de $D1$.

Demostración . Probemos en primer lugar que la sucesion generada por (2.38) está bien definida y contenida en X.
En efecto: Como f es continua en $\bar{X}$ es acotada, por lo cual podemos considerar a M como una cota inferior de f. Por lo cual:

$$f_k(x) = f(x) + D_h(x,x^k)/c_k \geq M + D_h(x,x^k)/c_k, \quad \forall x \in \bar{X}.$$

Por la propiedad (iii) de la función de Bregman tenemos que el conjunto de nivel asociado a D_h es acotado, además si $y \in \mathcal{L}_{f_k}(x,\alpha)$ se tiene que

$y \in \mathcal{L}_h(x, \alpha - M)$, de donde $\mathcal{L}_{f_k}(x, \alpha)$ es acotado y por tanto la minimización en (2.38) se reduce a un conjunto compacto, en consecuencia el mínimo es obtenido. Por otro lado como f es convexa, h es estrictamente convexa por la proposición 2.2.4,iii,(c) se tiene que D_h es estrictamente convexa y así f_k es estrictamente convexa, por tanto x^{k+1} es un minimizador único de f_k.
Ahora probaremos que $x^{k+1} \in X$.
Como x^{k+1} es el único minimizador de f_k, se deduce que x^{k+1} es el único $x \in \bar{X}$ que verifica:

$$0 \in \partial(f + D_h/c_k)(x).$$

Luego:

$$\nabla h(x^k)/c_k \in \partial(f + h/c_k)(x).$$

Si x está en la frontera de X, supongamos que existe $\xi \in \partial(f + h/c_k)(x)$.
Tomemos $z \in \bar{X}$ y definamos:

$$y^\ell = (1 - \epsilon_\ell)x + \epsilon_\ell z, \tag{2.39}$$

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

$$\lim_{\ell \to \infty} y^\ell = x.$$

Por (2.39) obtenemos:

$$\epsilon_\ell \xi^t(z - x) = \xi^t(y^\ell - x). \tag{2.40}$$

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

$$\xi^t(y^\ell - x) \leq f(y^\ell) - f(x) + (h(y^\ell) - h(x))/c_k. \tag{2.41}$$

Como h es estricta convexa y continuamente diferenciable, por la propiedades (i),(ii). Usamos la Proposición 1.1.4 para deducir:

$$f(y^\ell) - f(x) + (h(y^\ell) - h(x))/c_k \leq f(y^\ell) - f(x) + \nabla h(y^\ell)^t(y^\ell - x)/c_k. \tag{2.42}$$

Usando nuevamente (2.39) y la convexidad de f se obtiene:

$$f(y^\ell) - f(x) + \nabla h(y^\ell)^t(y^\ell - x)/c_k \leq \epsilon_\ell(f(z) - f(x)) + \frac{\epsilon_\ell}{c_k(1 - \epsilon_\ell)} \nabla h(y^\ell)^t(z - y^\ell). \tag{2.43}$$

Por lo cual de (2.40), (2.41), (2.42) y (2.43) se tiene:

$$\epsilon_\ell \xi^t(z-x) \leq \epsilon_\ell(f(z)-f(x)) + \frac{\epsilon_\ell}{c_k(1-\epsilon_\ell)} \nabla h(y^\ell)^t(z-y^\ell). \tag{2.44}$$

En consecuencia:

$$c_k(1-\epsilon_\ell)[f(x)-f(z)+\xi^t(z-x)] \leq \nabla h(y^\ell)^t(z-y^\ell). \tag{2.45}$$

Como $\lim\limits_{\ell\to\infty} y^\ell = x$ y además x está en la frontera de X, asimismo h es coerciva en la frontera, por lo cual el lado derecho de (2.45) tiende a $-\infty$ cuando $\ell \to \infty$, mientras el lado izquierdo tiene un límite finito, lo cual es contradictorio, por lo tanto $\partial(f + D_h) = \emptyset$, para todo x en la frontera de X, en consecuencia $x^{k+1} \in X$.

Probaremos ahora que: $D_h(\bar{x}, x^{k+1}) \leq D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k)$, para cada k y cada solución $\bar{x}$ del problema $D1$.

Usamos la Proposición 2.2.4(a), tomando $x = \bar{x},\ y = x^k,\ z = x^{k+1}$, obtenemos:

$$D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^{k+1} - \bar{x} \rangle. \tag{2.46}$$

Por (2.38) se deduce:

$$0 \in \partial[f + D_h(\cdot, x^k)/c_k](x^{k+1}). \tag{2.47}$$

Usando (2.47) y la Proposición 2.2.4(b).

$$[\nabla h(x^k) - \nabla h(x^{k+1})]/c_k \in \partial f(x^{k+1}). \tag{2.48}$$

Sea $y^k = [\nabla h(x^k) - \nabla h(x^{k+1})]/c_k$, por (2.46) y la definición de subgradiente:

$$\begin{aligned} D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) &= c_k \langle y^k, x^{k+1} - \bar{x} \rangle \\ &\geq c_k(f(x^{k+1}) - f(\bar{x})) \\ &\geq 0. \end{aligned} \tag{2.49}$$

Por tanto tenemos que se verifica:

$$D_h(\bar{x}, x^{k+1}) \leq D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k). \tag{2.50}$$

Probemos ahora que $\{x^k\}$ es acotada y $\lim_{k\to\infty} x^{jk} = \widehat{x}$ implica $\lim_{k\to\infty} x^{jk+1} = \widehat{x}$.
Por (2.50) se tiene que $\{D_h(\bar{x}, x^k)\}$ es decreciente y no negativa, y por tanto es convergente:

$$\lim_{k\to\infty} D_h(x^{k+1}, x^k) = 0. \tag{2.51}$$

Como $\{D_h(\bar{x}, x^k)\}$ es decreciente, se tiene $D_h(\bar{x}, x^k) \leq D_h(\bar{x}, x^0)$. Por tanto $\{x^k\}$ es acotada, en virtud que h es de Bregman (iii).
Si $\lim_{k\to\infty} x^{jk} = \widehat{x}$ para una subsucesión de $\{x^k\}$, entonces por (v) se deduce que $\lim_{k\to\infty} x^{jk+1} = \widehat{x}$.
Probemos que los puntos de acumulación de $\{x^k\}$ son solución del problema $D1$.

Tomemos una solución $\bar{x}$ de $D1$. Sea $\widehat{x}$ un punto de acumulación de $\{x^k\}$ y sea $\{x^{jk}\}$ una subsucesión de $\{x^k\}$ tal que $\lim_{k\to\infty} x^{jk} = \widehat{x}$. ya hemos probado que $\lim_{k\to\infty} x^{jk+1} = \widehat{x}$.
De (2.49) se deduce:

$$0 \leq c_k(f(x^{jk+1}) - f(\bar{x})) \leq D_h(\bar{x}, x^{jk}) - D_h(\bar{x}, x^{jk+1}) - D_h(x^{jk+1}, x^{jk}). \tag{2.52}$$

Como $D_h(\bar{x}, x^{jk}) - D_h(\bar{x}, x^{jk+1}) - D_h(x^{jk+1}, x^{jk}) \to 0$, cuando $k \to \infty$, obtenemos que $\lim_{k\to\infty} f(x^{jk+1}) = f(\bar{x})$, además se verifica (v), por lo cual se tiene que: $f(\widehat{x}) = f(\bar{x})$.
Como $\bar{X}$ es cerrado y $\{x^k\} \subset \bar{X}$, tenemos que $\widehat{x} \in \bar{X}$ y $\widehat{x}$ es una solución de $D1$.

Finalmente consideremos $\widehat{x}$ un punto de acumulación de la sucesión $\{x^k\}$ y tomamos una subsucesión $\{x^{jk}\}$ de $\{x^k\}$, tal que $\lim_{k\to} x^{jk} = \widehat{x}$, entonces por (iv)

$$\lim_{k\to\infty} D_h(\widehat{x}, x^{jk}) = 0.$$

Ya hemos demostrado que $\widehat{x}$ es una solución del problema $D1$ y por (2.50) se tiene que $\{D_h(\widehat{x}, x^k)\}$ es no negativa y decreciente que converge a 0, por lo tanto $\lim_{k\to} x^k = \widehat{x}$.

$\blacksquare$

Capítulo 3

MINIMIZACIÓN PROXIMAL CON FUNCIONES DE BREGMAN GENERALIZADAS

En este capítulo se estudian en detalles una nueva versión del método de punto proximal y otra de multiplicadores, ambas son propuestos por Kiwiel en [9]. Para ello se introduce una nueva versión de las funciones de Bregman, consideradas en [9], las cuales se denominan funciones de Bregman generalizadas. Con estas funciones se desarrollará en este capítulo el método de punto proximal con funciones de Bregman generalizadas (BPM). El propósito es estudiar la minimización proximal para funciones que no sean necesariamente diferenciables, para ello se usan los subgradientes en vez del gradiente y la función de Bregman h es sustituida por una función de Bregman generalizada.

De igual modo se presenta el método de multiplicadores inexactos, el cual resuelve el problema dual al resuelto por el BPM. Es interesante notar que Kiwiel considera los problemas de minimización, de manera inversa a como se consideran usualmente, es decir el primal que él toma tiene el formato del problema dual usual y el dual suyo es el primal que normalmente se considera.

Una serie de resultados expresados en varios teoremas son presentados por Kiwiel en su Paper [9], estos resultados expresan características interesantes de estos métodos, fundamentalmente su convergencia. En este capítulo se hace una demostración detallada de varios de esos teoremas. El estudio de estos resultados se justifica ya que alguno de ellos nos serán útiles para la implementación del método de multiplicadores basados en shifts con una penalidad no coerciva.

Además los métodos presentados por Kiwiel son semejantes a los presentados por Castillo en [3], de allí la importancia de estudiarlos, no obstante el aporte que se hace en este trabajo es analizar e implementar el método presentado por Castillo con una penalidad no coerciva, la cual no considera Kiwiel en sus análisis.

Comenzamos considerando el siguiente problema de minimización convexa:

$$f_* = \inf\{f(x) : x \in X\}, \tag{3.1}$$

donde $f : \mathbb{R}^n \to (-\infty, \infty]$ es una función cerrada, propia y convexa y X es un conjunto no vacío, cerrado y convexo en $\mathbb{R}^n$.

En el capítulo anterior vimos que el problema (3.1) puede resolverse usando el método de punto proximal generado por la sucesión:

$$x^{k+1} = \arg\min\{f(x) + \|x - x^k\|^2/2c_k \quad : x \in X\} \quad \text{para} \quad k = 1, 2, \ldots \quad . \tag{3.2}$$

Otra forma de resolver el problema (3.1) consiste en reemplazar $\|\cdot\|$ por la función D_h, de donde se deduce la siguiente sucesión:

$$x^{k+1} = \arg\min\{f(x) + D_h(x, x^k)/c_k \quad : x \in X\}, \tag{3.3}$$

donde $D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y\rangle$ se denomina la D-función de una función de Bregman h.

3-1 FUNCIONES DE BREGMAN GENERALIZADAS

En esta sección se introduce un concepto relacionado con las funciones de Bregman, el cual es una generalización del mismo, la ventaja de este nuevo concepto es que no se exige que la función sea necesariamente diferenciable.

Para una función h convexa, cerrada y propia en $\mathbb{R}^n$, posiblemente no diferenciable definimos las funciones distancias como sigue:

$$D_h^\flat(x,y) = h(x) - h(y) - \iota^*_{\partial h(y)}(x-y) \qquad \forall x,y \in \mathcal{D}_h. \tag{3.4}$$

$$D_h^\sharp(x,y) = h(x) - h(y) + \iota^*_{\partial h(y)}(y-x) \qquad \forall x,y \in \mathcal{D}_h. \tag{3.5}$$

Considerando que h es convexa y usando la Definición 1.1.13 se deduce:

$$h(x) \geq h(y) + \iota^*_{\partial h(y)}(x-y).$$

Y así $\quad D_h^\flat(x,y) \geq 0$
Además

$$D_h^\flat(x,y) \leq h(x) - h(y) - \langle \gamma, x-y \rangle \leq D_h^\sharp(x,y) \quad \forall x,y \in \mathcal{D}_h, \gamma \in \partial h(y).$$

Las funciones $D_h^\flat$ y $D_h^\sharp$ son generalizaciones de la D-función D_h usual definida por (2.36).

Por lo cual si la función h es diferenciable se obtiene:

$$D_h(x,y) = D_h^\flat(x,y) = D_h^\sharp(x,y) \quad \forall x \in D_h, y \in D_{\nabla h}.$$

Usando estas generalizaciones de D_h podemos generalizar también el concepto de funciones de Bregman de la siguiente forma.

Definición 3.1.1 FUNCIONES DE BREGMAN GENERALIZADAS *(ver [9], Definición 2.4)*
Sea $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ una función propia, convexa y cerrada, posiblemente no diferenciable, la función h es llamada función de Bregman Generalizada, la cual llamaremos B-función, si verifica lo siguiente:

Figura 3.1: $D_h^{\flat}(x,y)$

Figura 3.2: $D_h^{\sharp}(x,y)$

i) *h es estrictamente convexa en $\mathcal{D}_h$.*

ii) *h es continua en $\mathcal{D}_h$.*

iii) *Para cada $\alpha \in \mathbb{R}$ y $x \in \mathcal{D}_h$, el conjunto*

$$\mathcal{L}_h^{\flat}(x,\alpha) = \{y \in \mathcal{D}_{\partial h} : D_h^{\flat}(x,y) \leq \alpha\}$$

es acotado.

iv) *Para cada $\alpha \in \mathbb{R}$ y $x \in \mathcal{D}_h$, si $\{y^k\} \subset \mathcal{L}_h^{\flat}(x,\alpha)$ es un sucesión convergente con límite $y^* \in \mathcal{D}_h \backslash \{x\}$, entonces $D_h^{\sharp}(y^*, y^k) \to 0$.*

En este capítulo, salvo se indique lo contrario, cuando se haga refencia a las funciones de Bregman o B-función se considerará esta definición.

Por otro lado $D_h^\flat(x,y)$ y $D_h^\sharp(x,y)$ son usadas como casi-distancias ya que:

$$0 \leq D_h^\flat(x,y) \leq D_h^\sharp(x,y),$$

y

$$D_h^\flat(x,y) = 0 \Longleftrightarrow D_h^\sharp(x,y) = 0 \Longleftrightarrow x = y,$$

ya que h es estrictamente convexa.

El siguiente lema establece que la suma de funciones de Bregman y funciones poliedrales es una B-función:

Lema 3.1.1 *(ver [10], Lema 2.8)*
Sea $h = \sum\limits_{i=1}^{k} h_i$, *donde* $h_1, \ldots, h_k$ *son funciones cerradas, propias y convexas tal que:* $h_{j+1}, \ldots, h_k$ *son poliedrales y se cumple*

$$(\bigcap_{i=1}^{j} ri\ \mathcal{D}_{h_i}) \cap (\bigcap_{i=j+1}^{k} \mathcal{D}_{h_i}) \neq \emptyset,$$

sí además h_1 *es una B-función;* $h_2, \ldots, h_j$ *son continuas en* $\mathcal{D}_h = \bigcap\limits_{i=1}^{k} \mathcal{D}_{h_i}$ *y satisfacen la condición (iv) de la Definición 3.1.1, entonces* h *es una B-función. En particular* h *es B-función si las funciones* $h_1, \ldots, h_k$ *lo son.*

Demostración . Probemos cada una de las condiciones de la Definición 3.1.1

i) Como h_i es convexa para $i = 1, 2, \ldots, k$ y h_1 es estrictamente convexa, ya que es una B-función, se deduce que $h = \sum_{i=1}^{k} h_i$ es estrictamente convexa.

ii) Como h_i es cerrada, propia y convexa para: $i = 1, \ldots, k$. Usando (Corolario 7.5.1 y Teorema 10.1, [16]) se tiene que h_i es continua en D_{h_i} y así $h = \sum\limits_{i=1}^{k} h_i$ es continua en $\mathcal{D}_h = \bigcap\limits_{i=1}^{k} \mathcal{D}_{h_i}$.

iii) Como $h_{j+1},\ldots,h_k$ son poliedrales y $(\bigcap_{i=1}^{j} \operatorname{ri} \mathcal{D}_{h_i}) \cap (\bigcap_{i=j+1}^{k} \mathcal{D}_{h_i}) \neq \emptyset$ se deduce que $\partial h = \sum_{i=1}^{k} \partial h_i$ (ver [16], Teorema 23.8), por lo cual:

$$\begin{aligned}
D_h^\flat(x,y) &= h(x) - h(y) - \iota^*_{\partial h(y)}(x-y)\\
&= \sum_{i=1}^{k} h_i(x) - \sum_{i=1}^{k} h_i(y) - \iota^*_{\sum_{i=1}^{k}\partial h_i(y)}(x-y)\\
&= \sum_{i=1}^{k} h_i(x) - \sum_{i=1}^{k} h_i(y) - \sum_{i=1}^{k} \iota^*_{\partial h_i(y)}(x-y)\\
&= \sum_{i=1}^{k} D_{h_i}^\flat(x,y).
\end{aligned}$$

De donde si $y \in \mathcal{L}_h^\flat(x,\alpha)$ se deduce que:

$$\begin{aligned}
D_h^\flat(x,y) &\leq \alpha\\
\sum_{i=1}^{k} D_{h_i}^\flat(x,y) &\leq \alpha\\
D_{h_i}^\flat(x,y) &\leq \alpha \quad \forall i = 1,\ldots,k.
\end{aligned}$$

Por tanto $y \in \cap \mathcal{L}_{h_i}^\flat(x,\alpha)$, de donde $\mathcal{L}_h^\flat(x,\alpha) \subset \cap \mathcal{L}_{h_i}^\flat(x,\alpha)$.
Como $\mathcal{L}_{h_1}^\flat(x,\alpha)$ es acotado, ya que h_1 es B-función, deduciéndose que $\cap \mathcal{L}_{h_i}^\flat(x,\alpha)$ es acotado y por tanto $\mathcal{L}_h^\flat(x,\alpha)$ también es acotado.

iv) Tenemos en primer lugar que:

$$\begin{aligned}
D_h^\sharp(x,y) &= h(x) - h(y) + \iota^*_{\partial h(y)}(y-x)\\
&= \sum_{i=1}^{k} h_i(x) - \sum_{i=1}^{k} h_i(y) + \iota^*_{\sum_{i=1}^{k}\partial h_i(y)}(y-x)\\
&= \sum_{i=1}^{k} h_i(x) - \sum_{i=1}^{k} h_i(y) + \sum_{i=1}^{k} \iota^*_{\partial h_i(y)}(y-x)\\
&= \sum_{i=1}^{k} D_{h_i}^\sharp(x,y).
\end{aligned}$$

Consideremos una sucesión $\{y^k\} \subset \mathcal{L}_h^\flat(x,\alpha)$ convergente, con límite

$y^* \in \mathcal{D}_h \backslash \{x\}$.
Para $i = 1, \ldots, j$ se tiene por hipótesis $D^\sharp_{h_i}(y^*, y^k) \to 0$.
Para $i = j+1, \ldots, k$ consideremos S un subconjunto no vacío compacto de $\overset{\circ}{\mathcal{D}}_h$, usamos ([16], Teorema 24.7) para obtener:

$$|h_i(y) - h_i(x)| \leq \lambda |y - x| \quad \forall x, y \in S,$$

donde $\lambda = \sup\{|\gamma| : \gamma \in \partial h(S)\}$.
Como:

$$\begin{aligned} D^\sharp_{h_i}(y^*, y^k) &= h_i(y^*) - h_i(y^k) + \iota^*_{\partial h_i(y^k)}(y^k - y^*) \\ &\leq \lambda |y^k - y^*| + \iota^*_{\partial h_i(y^k)} |y^k - y^*| \\ &\leq 2\lambda |y^k - y^*|, \end{aligned} \tag{3.6}$$

se obtiene $D^\sharp_{h_i}(y^*, y^k) \to 0$, para $i = j+1, \ldots, k$, ya que $y^k \to y^*$, por lo tanto:

$$D^\sharp_h(y^*, y^k) = \sum_{i=1}^{j} D^\sharp_{h_i}(y^*, y^k) + \sum_{i=j+1}^{k} D^\sharp_{h_i}(y^*, y^k) \to 0.$$

∎

Podemos decir que la suma de k funciones de Bregman que tengan dominio común con $n-k$ funciones poliedrales con el mismo domino es una función de Bregman.

El lema siguiente presenta una caracterización para las B-función, en tal sentido se establece que la función sea esencialmente estricta convexa y el dominio de la conjugada sea abierto como condición necesaria y suficiente para que sea B-función.
También se establece que la suma de n funciones de Bregman con dominio común es una una B-función.

Lema 3.1.2 *(ver [10], Lemas 2.10 y 2.11)*

i) Sea h una función propia, cerrada y convexa en $\mathbb{R}$, entonces h es una B-función si sólo si h es esencialmente estricta convexa y $\mathcal{D}_{h^*} = \overset{\circ}{\mathcal{D}}_{h^*}$.

ii) Sean $h_1, \ldots, h_n$ B-funciones en $\mathbb{R}$, entonces $h = \sum_{i=1}^{k} h_i(x_i)$ es una B-función.

Demostración .

i) Si h es B-función entonces por definición es estrictamente convexa en $\mathcal{D}_h$ y por tanto es esencialmente estrictamente convexa.
Probemos ahora que $\mathcal{D}_{h^*} = \overset{\circ}{\mathcal{D}}_{h^*}$.
Supongamos que $\mathcal{D}_{h^*} \neq \overset{\circ}{\mathcal{D}}_{h^*}$. Tomamos $\bar{\gamma} = \max_{\gamma \in D_{h^*}} \gamma \in \mathcal{D}_{h^*}$.
Consideremos $\phi(z) = h(z) - \langle \bar{\gamma}, z \rangle + h^*(\bar{\gamma}) \geq 0$, por otro lado por (1.10) $\operatorname{im} \partial h \subset \mathcal{D}_{h^*}$.
Usando ([16], Teorema 23.8) se deduce:

$$\partial \phi(z) = \partial h(z) - \bar{\gamma} \subset -\mathbb{R}_+, \ \forall z,$$

por lo cual ϕ es no creciente y $[x, \infty) \subset \mathcal{D}_h, \quad \forall x \in \mathcal{D}_h$.
Existe $z^k \uparrow \infty, \quad \gamma^k \in \partial h(z^k), \quad \gamma^1 \leq \gamma^k \leq \bar{\gamma}$ y $x > 0$ tal que:

$$\begin{aligned}
h(z^k) + \langle \gamma^k, x - z^k \rangle &\geq \langle \bar{\gamma}, z^k \rangle - h^*(\bar{\gamma}) + \langle \gamma^k, x - z^k \rangle \\
&= -h^*(\bar{\gamma}) + \langle \bar{\gamma} - \gamma^k, z^k \rangle + \langle \gamma^k, x \rangle \\
&\geq -h^*(\bar{\gamma}) + \langle \gamma^k, x \rangle \\
&\geq -h^*(\bar{\gamma}) + \langle \gamma^1, x \rangle \\
&= h(x) - \alpha,
\end{aligned}$$

donde $\alpha = h^*(\bar{\gamma}) + h(x) + \langle \gamma^1, x \rangle < \infty$, por lo cual

$$h(x) - h(z^k) - \langle \gamma^k, x - z^k \rangle \leq \alpha,$$

y así

$$D_h^\flat(x, z^k) \leq \alpha.$$

Obteniéndose $\{z^k\} \subset \mathcal{L}_h^\flat(x, \alpha)$, por lo tanto $\mathcal{L}_h^\flat(x, \alpha)$ es no acotado, lo cual contradice el hecho de que h es B-función.
Similarmente se hace la prueba para $\bar{\gamma} = \min_{\gamma \in \mathcal{D}_{h^*}} \gamma \in \mathcal{D}_{h^*}$, por lo cual queda demostrado que $\mathcal{D}_{h^*} = \overset{\circ}{\mathcal{D}}_{h^*}$.

Recíprocamente, si h es cerrada, esencialmente estrictamente convexa y $\mathcal{D}_{h^*} = \overset{\circ}{\mathcal{D}}_{h^*}$, probemos que h es una B-función.
Como h es esencialmente estrictamente convexa se tiene que h es estrictamente convexa en $\mathcal{D}_{\partial h}$, pero como ri$\mathcal{D}_h \subset \mathcal{D}_{\partial h} \subset \mathcal{D}_h$, (ver Teorema 23.4, [16]) se tiene la convexidad estricta en $\mathcal{D}_h$.
La continuidad de h se puede establecer por ([16], Corolario 7.5.1 y Teorema 10.1).
Probemos que es válida la condición (iii) de la Definición 3.1.1, supongamos que $\mathcal{L}_h^\flat(x, \alpha/2)$ es no acotado para algún $x \in \overset{\circ}{\mathcal{D}}_h$ y $\alpha > 0$, entonces existe z^k y $\gamma^k \in \partial h(z^k)$ tal que:

$$h(z^k) + \langle \gamma^k, x - z^k \rangle \geq h(x) - \alpha, \quad |x - z^k| \geq 1 \quad \forall k \quad \text{y} \quad |z^k| \uparrow \infty.$$

Sea $\xi^k = x + (z^k - x)/|z^k - x|$, $\forall k$.
Como h es convexa, se obtiene

$$h(\xi^k) \leq \left(1 - \frac{1}{|z^k - x|}\right) h(x) + \frac{1}{|z^k - x|} h(z^k) < \infty,$$

por lo cual $\xi^k \in \mathcal{D}_h$ y

$$\begin{aligned} h(\xi^k) &\geq h(z^k) + \langle \gamma^k, \xi^k - z^k \rangle \\ &\geq h(x) - \alpha + \langle \gamma^k, \xi^k - x \rangle. \end{aligned}$$

Supongamos que $K \subset \{1, 2, \ldots\}$ es tal que $\{z^k\}_{k \in K} \uparrow \infty$, entonces $\{\gamma^k\}_{k \in K}$ es no decreciente, $\xi^k = x + 1$ y $\langle \gamma^k, \xi^k - x \rangle \leq h(\xi^k) - h(x) + \alpha$, $\forall k \in K$, y así $\{\gamma^k\}_{k \in K} \uparrow \gamma^\infty \in \mathbb{R}$, pero $h(z^k) + \langle \gamma^k, x - z^k \rangle \geq h(x) - \alpha$.
Por tanto $h^*(\gamma^k) \leq \alpha - h(x) + \langle \gamma^k, x \rangle$, de tal forma que si se toma límite $h^*(\gamma^\infty) < \infty$, ya que h^* es cerrada. Por lo cual $\gamma^\infty \in \mathcal{D}_{h^*} = \overset{\circ}{\mathcal{D}}_{h^*}$.
Como $\mathcal{D}_{h^*} = \overset{\circ}{\mathcal{D}}_{h^*}$ existe un $\bar{\gamma} \in \mathcal{D}_{h^*}$ tal que $\bar{\gamma} > \gamma^\infty$, entonces:

$$-h^*(\bar{\gamma}) + \langle \bar{\gamma}, z^k \rangle \leq h(z^k) \leq h(x) + \langle \gamma^k, z^k - x \rangle,$$

lo que implica que:

$$\langle \bar{\gamma} - \gamma^k, z^k \rangle \leq h(x) + h^*(\bar{\gamma}) - \langle \gamma^k, x \rangle < \infty, \quad \forall k.$$

Pero el lado izquierdo tiende a ∞ cuando $k \to \infty$, ya que $\{z^k\} \uparrow \infty$ y $\bar{\gamma} - \gamma^{\infty} > 0$. Obtenemos entonces una contradicción. Similarmente se obtiene una contradicción si $\{z^k\}_{k\in K} \downarrow -\infty$ usando $\{\gamma^k\}_{k\in K} \downarrow \gamma^{\infty} > \bar{\gamma} \in \overset{\circ}{\mathcal{D}}_{h^*}$. Por lo tanto $\mathcal{L}_h^\flat(x, \alpha/2)$ es acotado para todo $\alpha > 0$.

Probemos ahora que la condición (iv) es válida, en efecto:

Sea $\{y^k\} \in \mathcal{L}_h^\flat(x,\alpha)$ tal que $y^k \to y^*$.

Sea $l = \inf_{y\in D_h} y$ y sea $c = \sup_{y\in D_h} y$. Si $y^* \in (l, c)$, entonces por (3.6) tenemos $D_h^\sharp(y^*, y^k) \to 0$.

Si $y^k \downarrow y^* = l$, por la Proposición 1.1.7 se tiene:

$$\iota^*_{\partial h(y^k)}(1) = \sup_{s\in\partial h(y^k)} \{\langle 1, s\rangle\} = h'(y^k; 1) \downarrow h'(y^*; 1) \in [-\infty, \infty).$$

Si $h'(y^*; 1) > -\infty$, entonces por la continuidad de h y como $y^k \downarrow y^*$ se obtiene:

$$0 \leq D_h^\sharp(y^*, y^k) = h(y^k) - h(y^*) + (y^k - y^*)h'(y^k; 1) \to 0.$$

Si $h'(y^*; 1) = -\infty$ entonces $h'(y^k; 1) \leq 0$ para un k suficientemente grande, por lo tanto:

$$0 \leq D_h^\sharp(y^*, y^k) \leq h(y^k) - h(y^*) \to 0.$$

Para el caso $y^k \uparrow y^* = c$ lae prueba se realiza de forma análoga.

Se concluye que h es B-función.

ii) Sí $h_1, \ldots, h_n$ son B-funciones en $\mathbb{R}$, entonces son continuas y estrictamente convexas, por lo cual $h = \sum_{i=1}^{k} h_i(x_i)$ es continua y estrictamente convexa.

Sean $\alpha \in \mathbb{R}$, $x \in \bigcap_{i=1}^{k} \mathcal{D}_{h_i}$, consideremos a $y \in \mathcal{L}_h^\flat(x,\alpha)$, entonces $D_h^\flat(x,y) \leq \alpha$, pero $D_h^\flat(x,y) = \sum_{i=1}^{k} D_{h_i}^\flat(x,y)$, por lo cual $D_{h_i}^\flat(x,y) \leq \alpha$, para cada $i = 1, \ldots, k$; de donde $y \in \mathcal{L}_{h_i}^\flat(x,\alpha)$, para cada $i = 1, \ldots, k$. Por lo cual $\mathcal{L}_h^\flat(x,\alpha) \subset \bigcap_{i=1}^{k} \mathcal{L}_{h_i}^\flat(x,\alpha)$. Como cada $\mathcal{L}_{h_i}^\flat(x,\alpha)$ es acotado, por ser h_i una B-función, se deduce que $\mathcal{L}_h^\flat(x,\alpha)$ es acotado y se cumple la condición (iii).

Sean $\alpha \in \mathbb{R}$, $x \in \bigcap_{i=1}^{k} \mathcal{D}_{h_i}$, consideremos una sucesión $\{y^k\} \subset \mathcal{L}_h^\flat(x,\alpha)$ convergente a un $y^* \in \mathcal{D}_h \backslash \{x\}$. Tenemos que $\{y^k\} \subset \bigcap_{i=1}^{k} \mathcal{L}_{h_i}^\flat(x,\alpha)$, de donde $D_{h_i}^\sharp(y^*, y^k) \to 0$, para cada $i = 1, \ldots, k$. Por lo cual:

$$D_h^\sharp(y^*, y^k) = \sum_{i=1}^{k} D_{h_i}^\sharp(y^*, y^k) \to 0,$$

por lo tanto h es una B-función.

∎

Si una función es de Bregman, entonces su conjugada es esencialmente suave. Asimismo si una función propia cerrada y convexa es esencialmente suave y su dominio es abierto, entonces su conjugada es B-función. Estos resultados son establecidos en el siguiente lema.

Lema 3.1.3 *(ver [9], Lema 2.9)*

i) Si ψ es una B-función en $\mathbb{R}$, entonces ψ^* es esencialmente suave y $\mathcal{D}_{\psi^*} = \overset{\circ}{\mathcal{D}}_{\psi^*}$.

ii) Si $\phi : \mathbb{R} \to (-\infty, +\infty]$ es cerrada, propia, convexa y esencialmente suave y además $\mathcal{D}_\phi = \overset{\circ}{\mathcal{D}}_\phi$, entonces ϕ^* es una B-función con

$$ri\ \mathcal{D}_{\phi^*} \subset im\ \nabla\phi \subset \mathcal{D}_{\phi^*}.$$

Demostración .

i) Si ψ es B-función entonces por el Lema 3.1.2(i) es esencialmente estrictamente convexa y $\mathcal{D}_{\psi^*} = \overset{\circ}{\mathcal{D}}_{\psi^*}$. Por la Proposición 1.2.7(ii) ψ^* es esencialmente suave.

ii) Como ϕ es cerrada se tiene que $\phi^{**} = \phi$, además es propia y esencialmente suave; así por la Proposición 1.2.7(ii) se deduce que ϕ^* es esencialmente estrictamente convexa. Por lo tanto por el Lema 3.1.2(i) ϕ^* es B-función y se verifica $\mathcal{D}_\phi = \overset{\circ}{\mathcal{D}}_\phi$.

Por otro lado, usando (1.10) se tiene $\mathcal{D}_{\partial\phi^*} = \text{im } \partial\phi$, y por la Proposición 1.2.7(i) se deduce $\text{im } \partial\phi = \text{im } \nabla\phi$, por lo cual:

$$\text{ri}\mathcal{D}_{\phi^*} \subset \text{im}\nabla\phi \subset \mathcal{D}_{\phi^*}.$$

∎

Para relacionar la idea de distancias con la derivada direccional definimos a $D_h^{'}$ de la siguiente manera:

$$D_h^{'}(x,y) = h(x) - h(y) - h^{'}(y; x-y), \quad \forall x, y \in \mathcal{D}_h. \tag{3.7}$$

Notemos que $D_h^{'}(x,y) \geq 0$, ya que h es convexa. Además se verifica:

$$D_h^{'}(x,y) = h(x) - h(y) - h^{'}(y; x-y) \leq D_h^{\flat}(x,y), \quad \forall x, y \in \mathcal{D}_h.$$

En general:

$$0 \leq D_h^{'}(\cdot, x^k) \leq D_h^{\flat}(\cdot, x^k) \leq D_h^{k}(\cdot, x^k) \leq D_h^{\sharp}(\cdot, x^k) \tag{3.8}$$

Consideremos el conjunto $\mathcal{L}_h^{'}(x,\alpha) = \{y \in \mathcal{D}_h : D_h^{'}(x,y) \leq \alpha\}$. Tenemos que $\mathcal{L}_h^{\flat}(x,\alpha) \subset \mathcal{L}_h^{'}(x,\alpha)$, debido a que si $y \in \mathcal{L}_h^{\flat}(x,\alpha)$, entonces $D_h^{\flat}(x,y) \leq \alpha$, por lo cual $D_h^{'}(x,y) \leq \alpha$, de donde $y \in \mathcal{L}_h^{'}(x,\alpha)$.

Lema 3.1.4 *(ver [10], Lemas 2.15 y 2.16)*
Sea h una B-función en $\mathbb{R}^n$, entonces:

i) Si $\{x^k\}$ es una sucesión en $\mathcal{L}_h^{\flat}(x,\alpha)$ para algún $x \in \mathcal{D}_h, \alpha \in \mathbb{R}$, entonces $\{x^k\}$ es acotada y cada punto límite de $\{x^k\}$ está en $\mathcal{D}_h$.

ii) Si $\{x^k\} \subset \mathcal{D}_h$ es acotada y $D_h^{'}(x,x^k) \to 0$, para algún x, entonces $x^k \to x$.

Demostración .

i) Como $\{x^k\} \subset \mathcal{L}_h^{\flat}(x,\alpha)$, y por la Definición 3.1.1 $\mathcal{L}_h^{\flat}(x,\alpha)$ es acotado, se tiene que $\{x^k\}$ es acotada.
Sea y un punto límite de $\{x^k\}$, como $\{x^k\} \subset \mathcal{L}_h^{\flat}(x,\alpha)$, tenemos que

$$h(x) - h(x^k) - h^{'}(x^k; x - x^k) \leq \alpha, \quad \forall k,$$

luego, si usamos (Teorema 23.1,[16]), tomando $\lambda = 1$, tenemos:

$$h(x) - \alpha \leq h(x^k) + h^{'}(x^k; x - x^k) \leq h(x^k) + h(x^k + x - x^k) - h(x^k) = h(x), \quad \forall k.$$

Tomando límite cuando $x^k \to y$ se obtiene:

$$h(y) + h^{'}(y; x - y) \leq h(x) < \infty,$$

por lo cual $y \in \mathcal{D}_h$.

ii) Supongamos que $x^k \to x^*$, con $x^* \neq x$.
Usaremos el siguiente resultado expresado en ([10],Lema 2.14):

$$\limsup_{k \to \infty} h^{'}(x^k, y) \leq h^{'}(x^*, y), \quad \forall y \in \mathcal{D}_h. \tag{3.9}$$

Como $D_h^{'}(x, x^k) \to 0$, se tiene $h(x) - h(x^k) - h^{'}(x^k, x - x^k) \to 0$.
Por continuidad de h y de (3.9), al tomar límite obtenemos:

$$0 \geq h(x) - h(x^*) - h^{'}(x^*, x - x^*).$$

Lo cual contradice el hecho que h es estrictamente convexa, por tanto $x^* = x$.

∎

Podemos notar que se puede establecer una relación entre la suavidad esencial de una función y el hecho que su conjugada sea de Bregman.

EJEMPLOS DE FUNCIONES DE PENALIDAD

Para la aplicación de los métodos de multiplicadores se usan funciones que son llamadas de penalidad, usualmente se espera que está función sea esencialmente suave. De igual modo en los métodos de punto proximal se usan funciones de Bregman para su desarrollo, una de las funciones de Bregman que se pueden usar es la conjugada de la función de penalidad, estableciéndose así una importante relación de dualidad.

A continuación se presentan algunos ejemplos de funciones de penalidad con sus correspondientes conjugadas, tales ejemplos los considera Kiwiel en [9], además se presenta la función de penalidad que se usará para aplicar el método de multiplicadores basados en shifts, la cual considera Zang en [19].

1. Si $h(x) = \sum_{i=1}^{n} h_i(x_i)$, con $h_i : \mathbb{R} \to (-\infty, \infty]$, entonces $h^*(x) = \sum_{i=1}^{n} h_i^*(x_i)$ y $D_h(x, y) = \sum_{i=1}^{n} D_{h_i}(x_i, y_i)$.
2. $h(x) = |x|^\alpha/\alpha$ en $\mathcal{D}_h = \mathbb{R}$, para $\alpha > 1$. Entonces $h^*(\cdot) = |\cdot|^\beta/\beta$ con $\alpha + \beta = \alpha\beta$.
   Sí $\alpha = 2$ $h(x) = |x|^2/2$, tenemos que $h^*(\cdot) = |\cdot|^2/2$ y $D_h(x, y) = |x - y|^2/2$.
3. $h(x) = -x^\alpha/\alpha$ en $\mathcal{D}_h = \mathbb{R}_+$, con $\alpha \in (0, 1)$. Entonces $h^*(y) = -(-y)^\beta/\beta$ en $\mathcal{D}_h = (-\infty, 0)$ con $\alpha + \beta = \alpha\beta$.
4. $h(x) = \frac{\alpha x - x^\alpha}{1-\alpha}$ en $\mathcal{D}_h = \mathbb{R}_+$, con $\alpha \in (0, 1)$, entonces $h^*(y) = (1+y/\beta)^\beta$ en $\mathcal{D}_{h^*} = (-\infty, -\beta)$. Para $\alpha = 1/2$ se tiene $D_h(x, y) = (x^{1/2} - y^{1/2})^2/y^{1/2}$.
5. Penalidad $xlogx$:
   $h(x) = x \ln x$ en $\mathcal{D}_h = R_+(0 \ln 0 = 0)$, entonces $h^*(y) = \exp(y - 1)$ en $\mathcal{D}_{h^*} = \mathbb{R}$ y $D_h(x, y) = x \ln(x/y) + y - x$. Para $h(x) = x \ln x - x$ se tiene $h^*(y) = \exp y$.
6. Penalidad de Burg:
   $h(x) = -\ln x$ en $\mathcal{D}_h = \mathbb{R}_>$ se tiene $h^*(y) = -\ln(-y) - 1$ en $\mathcal{D}_{h^*} = (-\infty, 0)$ y $D_h(x, y) = -\ln(x/y) + x/y - 1$.
7. Penalidad de Helliger:
   $h(x) = -\sqrt{1 - x^2}$ en $\mathcal{D}_h = [-1, 1]$, entonces $h^*(y) = \sqrt{1 + y^2}$ en $\mathcal{D}_{h^*} = \mathbb{R}$ y $D_h(x, y) = \frac{1-xy}{\sqrt{1-y^2}} - \sqrt{1 - x^2}$ en $[-1, 1] \times (-1, 1)$.
8. $h(x) = -\sqrt{x(1 - x)}$ en $\mathcal{D}_h = [0, 1]$, entonces $h^*(y) = \frac{1}{2}(\ y+\sqrt{1 + y^2}\ )$.
9. Penalidad de Fermic-Dirac:
   $h(x) = x \ln x + (1 - x) \ln(1 - x)$ en $\mathcal{D}_h = [0, 1]$. Entonces $h^*(y) = \ln(1 + \exp y)$ en $D_{h^*} = \mathbb{R}$.

10. $h(x) = -\ln x - \ln(1-x)$ en $\mathcal{D}_h = [0,1]$, entonces
$h^*(y) = \frac{1}{2}[(y^2+4)^{1/2} + y - 2] + \ln[(y^2+4)^{1/2} - 2] - \ln y^2$ para $y \neq 0$
y $h^*(0) = -\ln 4$, $\mathcal{D}_{h^*} = \mathbb{R}$.

11. Consideremos la función, presentada por Zang en [19]:
Esta función es la que será usada para aplicar el método de multiplicadores basados en shifts con una penalidad no coerciva y será estudiada a profundidad en el capítulo 4 del presente trabajo.

$$\theta_r(x) = \begin{cases} 0 & \text{si} \quad x \leq -r \\ b_r(x) & \text{si} \quad -r \leq x \leq r \\ x & \text{si} \quad x \geq r, \end{cases} \tag{3.10}$$

donde $r > 0$ y $b_r(x) = \frac{1}{4r}x^2 + \frac{1}{2}x + \frac{1}{4}r$.
Podemos obtener:

$$\begin{aligned} b_r(x) &= \frac{1}{4r}x^2 + \frac{1}{2}x + \frac{1}{4}r \\ &= r\left(\frac{x^2}{4r^2} + \frac{x}{2r} + \frac{1}{4}\right) \\ &= r\left(b\left(\frac{x}{r}\right)\right). \end{aligned} \tag{3.11}$$

Donde:

$$\begin{aligned} b(x) &= \frac{x^2}{4} + \frac{1}{2}x + \frac{1}{4} \\ &= \frac{1}{4}\left(x^2 + 2x + 1\right) \\ &= \frac{1}{4}\left(x+1\right)^2, \end{aligned} \tag{3.12}$$

por lo cual podemos definir:

$$\theta(x) = \begin{cases} 0 & \text{si} \quad x \leq -1 \\ b(x) & \text{si} \quad -1 \leq x \leq 1 \\ x & \text{si} \quad x \geq 1, \end{cases} \tag{3.13}$$

donde $b(x) = \frac{1}{4}\left(x+1\right)^2$
Notemos que $\theta(x) = r\theta_r(\frac{x}{r})$.
Calculemos ahora la conjugada θ^* de la función θ.

Calculemos en primer lugar $b^*(y) = \sup_x \{xy - b(x)\}$.
Sea $g(x) = xy - b(x)$, luego $g'(x) = y - \frac{1}{2}(x+1)$.
Como $g'(x) = 0$, se tiene $x = 2y - 1$,
y así

$$\begin{aligned} b^*(y) &= y(2y-1) - \frac{1}{4}(2y-1+1)^2 \\ &= 2y^2 - y - y^2 \\ &= y(y-1) \end{aligned} \tag{3.14}$$

Por otro lado como $-1 \leq x \leq 1$ y $x = 2y - 1$, entonces $0 \leq y \leq 1$, por lo cual $\mathcal{D}_{b^*} = [0,1]$.
Sí $x < -1$, entonces $h(x) = 0$, por lo cual consideramos la función

$$g(x) = xy,$$

de donde $g^{'}(x) = y$, pero como $g^{'}(x) = 0$, se tiene que $y = 0$, deduciéndose de esta forma que si $x < -1$, sólo se puede calcular $\theta^*(0) = 0$.
Análogamente se prueba que si $x > 1$, sólo se puede calcular $\theta^*(1) = 1$.
Podemos concluir que:

$$\theta^*(y) = y(y-1), \tag{3.15}$$

y $\mathcal{D}_{\theta^*} = [0,1]$.

3-2 MÉTODO DE PUNTO PROXIMAL CON FUNCIÓN DE BREGMAN GENERALIZADA

Una versión más general del método de punto proximal es presentado por Kiwiel en [9], la cuál denomina Minimización Proximal de Bregman (BPM), tal método usa en su desarrollo una función de Bregman generalizada, la cual no es necesariamente diferenciable.

3-2-1 ALGORITMO DEL BPM

Para aplicar el Método de Punto Proximal con funciones de Bregman Generalizada se considera en primer lugar el problema resolver:

$$(P1) \qquad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & x \in X \subset \mathbb{R}^n. \end{array}$$

Cumpliendo las siguientes hipótesis:

Hipótesis 3.2.1

i) f *es una función cerrada, propia y convexa.*

ii) X *es un conjunto no vacío, cerrado y convexo.*

iii) h *es una B-función (posiblemente no diferenciable).*

iv) $\mathcal{D}_{f_X} \bigcap \mathcal{D}_h \neq \emptyset$, *donde* $f_X = f + \iota_X$ *es la función objetivo esencial del problema (P1).*

v) $\{c_k\}$ *es una sucesión de números positivos que sastifacen* $\sum_{k=1}^{\infty} c_k = \infty$.

vi) $\{\epsilon_k\}$ *es una sucesión de números no negativos que sastifacen:*

$$\lim_{l\to\infty} \sum_{k=1}^{l} c_k\epsilon_k / \sum_{k=1}^{l} c_k = 0,$$

es decir: $\sum_{k=1}^{\infty} c_k\epsilon_k < \infty$.

El BPM lo podemos expresar a través del siguiente algoritmo:
Para la iteración $k \geq 1$, tenemos lo siguiente:

$$x^k \in \mathcal{D}_{f_X} \cap \mathcal{D}_{\partial h}. \tag{3.16}$$

$$\gamma^k \in \partial h(x^k). \tag{3.17}$$

$$D_h^k(x, x^k) = h(x) - h(x^k) - \langle \gamma^k, x - x^k \rangle \qquad \forall x. \tag{3.18}$$

Consideremos $\phi_k(x) = f_x(x) + D_h^k(x, x^k)/c_k$.
Hallar x^{k+1}, γ^{k+1} y p^{k+1} que satisfagan:

$$\gamma^{k+1} \in \partial h(x^{k+1}), \tag{3.19}$$

$$c_k p^{k+1} + \gamma^{k+1} - \gamma^k = 0, \tag{3.20}$$

$$p^{k+1} \in \partial_{\epsilon_k} f_X(x^{k+1}), \tag{3.21}$$

$$f_X(x^{k+1}) \leq f_X(x^k). \tag{3.22}$$

Debemos notar que: x^{k+1} es un ϵ_k-minimizador de $\phi_k(x) = f_X(x) + D_h^k(x, x^k)/c_k$, donde por (3.8) tenemos:

$$0 \leq D_h^\flat(\cdot, x^k) \leq D_h^k(\cdot, x^k) \leq D_h^\sharp(\cdot, x^k).$$

Una de las ventajas de este método es que no se requiere que la función h sea diferenciable, ya que en vez de usar el gradiente de h se usan subgradientes. Además al aplicar el método se obtiene la función objetivo ϕ_k, la cual es estrictamente convexa, por lo que se puede asegurar que el minimizador que se obtenga es único, esto garantiza que la sucesión $\{x^k\}$ esté bien definida.

A continuación se introduce una definición donde se considera la compacidad, pero con relación a una función. Estableciéndose a través de la acotación de todos los conjuntos de nivel.

Definición 3.2.1 *Una función ϕ cerrada, propia y estrictamente convexa en $\mathbb{R}^n$ es inf-compacta si el conjunto de nivel:*

$$\mathcal{L}_\phi(\alpha) = \{x : \phi(x) \leq \alpha\},$$

es acotado para cada $\alpha \in \mathbb{R}$.

Debemos observar que si $\mathcal{L}_\phi(\alpha) = \{x : \phi(x) \leq \alpha\}$ es acotado para algún $\alpha \in \mathbb{R}$, entonces es acotado para cada $\alpha \in \mathbb{R}$.

En efecto, supongamos que $\mathcal{L}_\phi(\alpha^*) = \{x : \phi(x) \leq \alpha^*\}$ es acotado para un $\alpha^* \in \mathbb{R}$. Tomemos un $\alpha \in \mathbb{R}$ arbitrario.
Si $\alpha \leq \alpha^*$, entonces $\mathcal{L}_\phi(\alpha) \subset \mathcal{L}_\phi(\alpha^*)$, por lo cual $\mathcal{L}_\phi(\alpha)$ es acotado.
Si $\alpha > \alpha^*$, supongamos que existe un $y \in \mathbb{R}^n$ distinto de cero que pertenece al cono de recesión de $\mathcal{L}_\phi(\alpha)$, en virtud de lo cual si $x \in \mathcal{L}_\phi(\alpha)$, usando la Definición 1.1.19 se tiene que $x + \lambda y \in \mathcal{L}_\phi(\alpha)$, con $\lambda \geq 0$. Por lo tanto

$\phi(x+\lambda y) \leq \alpha$.
Como ϕ es estrictamente convexa, ya que f_X es convexa y h es estrictamente convexa, por ser B-función; tenemos que existe un único x^* tal que:

$$\phi(x^*) \leq \phi(x), \quad \forall x \in \mathcal{L}_\phi(\alpha),$$

por lo cual:

$$\phi(x^*) \leq \phi(x^*+\lambda y), \quad \forall \lambda \geq 0.$$

Además si $\lambda_1 < \lambda_2$, entonces $\phi(x^*+\lambda_1 y) < \phi(x^*+\lambda_2 y)$, luego existe un $\lambda \geq 0$ tal que $\phi(x^*+\lambda y) > \alpha$, lo cual es contradictorio. De donde el cono de recesión de $\mathcal{L}_\phi(\alpha)$ es el conjunto $\{0\}$, por lo cual usando ([6], Proposición 2.2.3) se obtiene que $\mathcal{L}_\phi(\alpha)$ es acotado.

Con este resultado podemos decir que la función ϕ es inf-compacta si $\mathcal{L}_\phi(\alpha)$ es acotado para algún $\alpha \in \mathbb{R}$.

El lema siguiente, permite determinar la existencia de un único minimizador de una función a través del concepto de inf-compacidad.

Lema 3.2.1 *(Ver [9],Lema 3.2)*
Una función ϕ cerrada, propia y estrictamente convexa en $\mathbb{R}^n$ tiene un único minimizador si sólo si ϕ es inf-compacta.

Demostración . Si $x \in \arg\min \phi$, entonces por convexidad estricta de ϕ se tiene que $\mathcal{L}_\phi(\phi(x)) = \{x\}$ es acotado y así ϕ es inf-compacta.
Recíprocamente si ϕ es inf-compacta se tiene que para algún $\alpha \in \mathbb{R}$ el conjunto $\mathcal{L}_\phi(\alpha) \neq \emptyset$ es acotado, entonces es cerrado y contiene a $\arg\min \phi \neq \emptyset$ ya que ϕ es cerrada. ∎

La sucesión $\{x^k\}$ del BPM es generada por los minimizadores de la función ϕ_k, de esta forma el resultado anterior garantiza que la sucesión $\{x^k\}$ está bien definida si ϕ_k es inf-compacta.

El siguiente lema establece que la función ϕ_k es cerrada, propia y estrictamente convexa; además cada x^{k+1} de la sucesión generada por el BPM es un ϵ_k- minimizador de ϕ_k. También se establecen condiciones para que la función ϕ_k sea inf-compacta y por consiguiente la sucesión $\{x^k\}$ esté bien definida.

Lema 3.2.2 *(Ver [9], Lema 3.3)*
Con las hipótesis (3.2.1) del problema (P1) se tiene lo siguiente:

i) ϕ_k *es cerrada, propia y estrictamente convexa.*

ii) $\phi_k(x^{k+1}) \leq \inf \phi_k + \epsilon_k$, *es decir* $0 \in \partial_{\epsilon_k}\phi_k(x^{k+1})$.

iii) Si $f_* = \inf_X f > -\infty$, *entonces* ϕ_k *es inf-compacta.*

iv) ϕ_k *es inf-compacta si* $(\gamma^k - c_k im\, \partial f_X) \cap im\, \partial h \neq \emptyset$, *donde* $im\, \partial h = \overset{\circ}{\mathcal{D}}_{h^*}$, *de modo que* $im \partial h = \mathbb{R}^n$ *si sólo si* h *es cofinita. En particular,* ϕ_k *es inf-compacta si* $(\gamma^k - c_k ri\, \mathcal{D}_{f_X^*}) \cap ri\, \mathcal{D}_{h^*} \neq \emptyset$.

v) Si ϕ_k *es inf-compacta y se cumple alguna de estas propiedades:*

$$ri\mathcal{D}_{f_X} \cap ri\, \mathcal{D}_h \neq \emptyset \quad \text{ó} \quad \mathcal{D}_{f_X} \cap ri\, \mathcal{D}_h \neq \emptyset,$$

además si f_X *es poliédrica, entonces existen:*

$$\widehat{x}^{k+1} = \arg\min \phi_k, \quad \widehat{p}^{k+1} \in \partial f_X(\widehat{x}^{k+1}) \ y \quad \widehat{\gamma}^{k+1} \in \partial h(\widehat{x}^{k+1}),$$

tal que:

$$f_X(\widehat{x}^{k+1}) + D_h^k(\widehat{x}^{k+1}, x^k)/c_k \leq f_X(x^k)$$

y $c_k\widehat{p}^{k+1} + \widehat{\gamma}^{k+1} - \gamma^k = 0$.
También $\widehat{x}^{k+1} \in \overset{\circ}{\mathcal{D}}_h$ *si* $\mathcal{D}_{f_X} \subset \overset{\circ}{\mathcal{D}}_h$ *ó* $\mathcal{D}_{\partial h} = \overset{\circ}{\mathcal{D}}_h$; *es decir* h *es esencialmente suave.*

vi) Las hipotesís de (v) son válidas si se cumple algunas de las dos condiciones: ($ri\, \mathcal{D}_{\partial h} \subset \overset{\circ}{\mathcal{D}}_h$ *e* $\inf_X f > -\infty$) *ó* ($\mathcal{D}_{\partial f_X} \subset \overset{\circ}{\mathcal{D}}_h$ *e* $im\, \partial h = \mathbb{R}^n$).

Demostración .

i) Como f, ι_X, y h son cerradas propias y convexas; además los conjuntos $\mathcal{D}_{f_X} \cap X$, $\mathcal{D}_h$ y $\mathcal{D}_{f_X} \cap \mathcal{D}_h$ son no vacíos, entonces: $f_X = f + \iota_X$, $D_h^k(\cdot, x^k)$ y

$\phi_k = f_X + D_h^k(\cdot, x^k)/c_k$ son cerradas, propias y convexas y la convexidad estricta de ϕ_k se obtiene por que h lo es al ser una función de Bregman.

ii) Sea x arbitrario, usando (3.17); (3.18); (3.19) se tiene:

$$\begin{aligned}
D_h^k(x, x^k) &= h(x) - h(x^k) - \langle \gamma^k, x - x^k \rangle \\
&= h(x) - h(x^k) - \langle \gamma^k, x - x^k \rangle + h(x^{k+1}) - \langle \gamma^k, x^{k+1} - x^k \rangle - \\
&\quad - h(x^{k+1}) + \langle \gamma^k, x^{k+1} - x^k \rangle \\
&= D_h^k(x^{k+1}, x^k) + h(x) - h(x^{k+1}) + \langle \gamma^k, x^{k+1} - x \rangle \\
&\geq D_h^k(x^{k+1}, x^k) + \langle \gamma^{k+1}, x - x^{k+1} \rangle + \langle \gamma^k, x^{k+1} - x \rangle \\
&= D_h^k(x^{k+1}, x^k) + \langle \gamma^{k+1} - \gamma^k, x - x^{k+1} \rangle.
\end{aligned}$$

Dividiendo por c_k se obtiene:

$$D_h^k(x, x^k)/c_k \geq D_h^k(x^{k+1}, x^k)/c_k + \langle \gamma^{k+1} - \gamma^k, x - x^{k+1} \rangle / c_k.$$

Usando (3.20) se deduce:

$$D_h^k(x, x^k)/c_k \geq D_h^k(x^{k+1}, x^k)/c_k - \langle p^{k+1}, x - x^{k+1} \rangle. \tag{3.23}$$

Por (3.21) se tiene:

$$f_X(x) \geq f_X(x^{k+1}) + \langle p^{k+1}, x - x^{k+1} \rangle - \epsilon_k. \tag{3.24}$$

Así de (3.23) y (3.24) obtenemos:

$$f_X(x) + D_h^k(x, x^k)/c_k \geq f_X(x^{k+1}) + D_h^k(x^{k+1}, x^k)/c_k - \epsilon_k,$$

por lo cual

$$\phi_k(x) \geq \phi_k(x^{k+1}) - \epsilon_k.$$

$$\phi_k(x) + \epsilon_k \geq \phi_k(x^{k+1}).$$

Como x es arbitrario se tiene $\phi_k(x^{k+1}) \leq \inf \phi_k(x) + \epsilon_k$.

iii) Sea $\psi = D_h^k(\cdot, x^k)$, la cual es cerrada, propia y estrictamente convexa. Además $D_h^k(\cdot, x^k) \geq 0$, por lo cual se obtiene que:

$$\mathcal{L}_\psi(0) = \{x : \psi(x) \leq 0\} = \{x^k\}, \tag{3.25}$$

por tanto si usamos el Lema 3.2.1 se deduce que ψ es inf-compacta.
Sea $\beta = \inf \phi_k$, como $f_* \leq f_X(x)$, $\forall x \in \mathbb{R}^n$, además $\psi \geq 0$, por lo cual se deduce que $\beta \geq f_*$.

$$
\begin{aligned}
\emptyset \neq \mathcal{L}_{\phi_k}(\beta+1) &= \{x : \phi_k(x) \leq \beta+1\} \\
&= \{x : f_X(x) + \psi(x)/c_k \leq \beta+1\} \\
&= \{x : \psi(x) \leq c_k(\beta+1-f_X(x))\} \subset \mathcal{L}_{\psi}(c_k(\beta+1-f_*)).
\end{aligned}
$$

Por (3.25) y usando el Lema 3.2.1 se deduce que $\mathcal{L}_{\psi}(c_k(\beta+1-f_*))$ es acotado, por lo cual $\mathcal{L}_{\phi_k}(\beta+1)$ también es acotado y así ϕ_k es inf-compacta.

iv) Como $(\gamma^k - c_k \mathrm{im}\ \partial f_X) \cap \mathrm{im}\ \partial h \neq \emptyset$, podemos elegir $\hat{y} \in \mathcal{D}_{\partial f_X}$, $\hat{\gamma} \in \partial f_X(\hat{y})$, $\tilde{x} \in \mathcal{D}_{\partial h}$ y $\tilde{\gamma} \in \partial h(\tilde{x})$ tal que satisfagan:

$$\gamma^k - c_k\hat{\gamma} = \tilde{\gamma}.$$

Entonces $\tilde{\psi}(\cdot) = f_X(\hat{y}) + \langle \hat{\gamma}, \cdot - \hat{y}\rangle + D_h^k(\ \cdot, x^k)/c_k$ es cerrada, propia y estrictamente convexa.
Si consideramos que:

$$\partial\tilde{\psi}(\tilde{x}) = \hat{\gamma} + (\partial h(\tilde{x}) - \gamma^k)/c_k$$

Además $\tilde{\gamma} \in \partial h(\tilde{x})$ y $\gamma^k - c_k\hat{\gamma} = \tilde{\gamma}$, por lo cual:

$$0 \in \partial\tilde{\psi}(\tilde{x}),$$

por consiguiente

$$\tilde{x} = \arg\min \tilde{\psi}.$$

Por el Lema 2.20 $\tilde{\psi}$ es inf-compacta.
Consideremos un $\alpha \in \mathbb{R}$, entonces $\mathcal{L}_{\tilde{\psi}}(\alpha)$ es acotado.
Sea $x \in \mathcal{L}_{\phi_k}(\alpha)$, luego $\phi_k(x) = f_X(x) + D_h^k(x, x^k)/c_k \leq \alpha$.
Como $\hat{\gamma} \in \partial f_X(\hat{y})$, se tiene:

$$f_X(\hat{y}) + \langle \hat{\gamma}, x - \hat{y}\rangle \leq f_X(x),$$

por lo cual $\tilde{\psi}(x) \leq \phi(x) \leq \alpha$, de donde $x \in \mathcal{L}_{\tilde{\psi}}(\alpha)$, lo cual significa que $\mathcal{L}_{\phi_k}(\alpha) \subset \mathcal{L}_{\tilde{\psi}}(\alpha)$, deduciéndose que $\mathcal{L}_{\phi_k}(\alpha)$ es acotado y por tanto ϕ_k es inf-compacta.

Por convexidad estricta de h y usando (1.10) se deduce:

$$\mathrm{im}\partial h = \mathcal{D}_{\partial h^*} = \overset{\circ}{\mathcal{D}}_{h^*} \, .$$

Por otra parte $\overset{\circ}{D}_{h^*} = \mathbb{R}^n$ si sólo si $\mathcal{D}_{h^*} = \mathbb{R}^n$ si sólo si h es cofinita. (ver Proposición 1.2.3)

De manera particular ϕ_k es inf-compacta si $(\gamma^k - c_k \mathrm{ri}\ \mathcal{D}_{f_X^*}) \cap \mathrm{ri}\, \mathcal{D}_{h^*} \neq \emptyset$, ya que $\mathrm{ri}\mathcal{D}_{f_X^*} \subset \mathcal{D}_{\partial f_X^*} = \mathrm{im}\partial f_X$.

v) Por la parte (i) y el Lema 3.2.1 $\hat{x}^{k+1} = \arg\min \phi_k$ está bien definido. Luego

$$0 \in \partial\phi_k(\hat{x}^{k+1}) = \partial f(\hat{x}^{k+1}) + c_k(\partial h(\hat{x}^{k+1}) - \gamma^k).$$

Entonces existen $\hat{\gamma}^{k+1} \in \partial h(\hat{x}^{k+1})$ y $\hat{p}^{k+1} \in \partial f_X(\hat{x}^{k+1})$ tal que:

$$c_k \hat{p}^{k+1} + \hat{\gamma}^{k+1} - \gamma^k = 0.$$

Además de $\hat{x}^{k+1} = \arg\min \phi_k$ se deduce:

$$f_X(\hat{x}^{k+1}) + D_h^k(\hat{x}^{k+1}, x^k)/c_k \leq f_X(x^k) + D_h^k(x^k, x^k)/c_k = f_X(x^k).$$

Por otro lado $\hat{x}^{k+1} \in \overset{\circ}{\mathcal{D}}_h$, ya que $\hat{x}^{k+1} \in \mathcal{D}_{\partial f_X}$ y $\hat{x}^{k+1} \in \mathcal{D}_{\partial h}$, y por hipótesis $\mathcal{D}_{\partial f_X} \subset \overset{\circ}{\mathcal{D}}_h$ ó $\mathcal{D}_{\partial h} = \overset{\circ}{\mathcal{D}}_h$.

Finalmente por la Proposición 1.2.7 h es esencialmente suave.

vi) Si $\inf_X f > -\infty$ por la parte (iii) ϕ_k es inf-compacto. Si $\mathrm{im}\partial h = \mathbb{R}^n$ por la parte (iv) ϕ_k es inf-compacto. Por otro lado si $\mathrm{ri}\mathcal{D}_{f_X} \subset \overset{\circ}{D}_h$ entonces $\mathrm{ri}\mathcal{D}_{f_X} \cap \mathrm{ri}\mathcal{D}_h = \mathrm{ri}\mathcal{D}_{f_X} \neq \emptyset$.

$\blacksquare$

Este lema expresa resultados importantes, en primer lugar provee unas características deseables de la función ϕ_k, determinando que es propia, cerrada y estrictamente convexa. Como la sucesión $\{x^{k+1}\}$ consiste en el ϵ_k-minimizador de la función ϕ_k, se tiene que la sucesión está bien definida por la convexidad estricta de ϕ_k.

Por otro lado se establecen las condiciones para que la función ϕ_k sea inf-compacta, una de las condiciones consiste en que el ínfimo de f en X sea mayor que $-\infty$.

3-2-2 CONVERGENCIA DEL BPM-

En cualquier método de minimización que genere una sucesión, es deseable que se pueda asegurar la convergencia de la mencionada sucesión, no obstante no siempre es posible asegurarlo, en algunos casos es necesario imponer algunas condiciones para obtener la convergencia.

Serán presentados a continuación varios lemas y teoremas, considerados por Kiwiel en [9], los cuales hacen un estudio pormenorizado de la convergencia de la sucesíón $\{x^k\}$ generada por el BPM.

Derivemos en primer lugar un radio de convergencia global del BPM.
Sea $s_k = \sum_{j=1}^{k} c_j$ para todo k.

Lema 3.2.3 *(Ver [9], Lema 4.1)*
$\forall x \in D_h$ y $k \geq l$ tenemos:

i)

$$\begin{aligned} D_h^{k+1}(x, x^{k+1}) + D_h^k(x^{k+1}, x^k) - D_h^k(x, x^k) &= \langle \gamma^k - \gamma^{k+1}, x - x^{k+1} \rangle \quad (3.26) \\ &\leq c_k[f_X(x) - f_X(x^{k+1})] + c_k\epsilon_k. \end{aligned}$$

ii) Si $f_X(x) \leq f_X(x^{k+1})$

$$D_h^{k+1}(x, x^{k+1}) \leq D_h^k(x, x^k) - D_h^k(x^{k+1}, x^k) + c_k\epsilon_k. \tag{3.27}$$

iii)

$$\begin{aligned} s_l[f_X(x^{l+1}) - f_X(x)] &\leq D_h^1(x, x^1) - D_h^{l+1}(x, x^{l+1}) - \sum_{k=1}^{l} D_h^k(x^{k+1}, x^k) + \\ &+ \sum_{k=1}^{l} c_k\epsilon_k. \end{aligned} \tag{3.28}$$

iv)

$$f_X(x^{l+1}) - f_X(x) \le D_h^1(x, x^1)/s_l + \sum_{k=1}^{l} c_k \epsilon_k / s_l. \qquad (3.29)$$

Demostración .

i) Sea $x \in \mathcal{D}_h$ arbitrario. Como $p^{k+1} \in \partial_{\epsilon_k} f_X(x^{k+1})$, entonces se tiene:

$$f_X(x) - f_X(x^{k+1}) + \epsilon_k \ge \langle p^{k+1}, x - x^{k+1} \rangle.$$

Como $\gamma^k - \gamma^{k+1} = c_k p^{k+1}$ y $c_k > 0$, resulta:

$$c_k[f_X(x) - f_X(x^{k+1})] + c_k \epsilon_k \ge \langle \gamma^k - \gamma^{k+1}, x - x^{k+1} \rangle. \qquad (3.30)$$

Tomando

$$I = D_h^{k+1}(x, x^{k+1}) + D_h^k(x^{k+1}, x^k) - D_h^k(x, x^k), \qquad (3.31)$$

se tiene:

$$\begin{aligned} I &= h(x) - h(x^{k+1}) - \langle \gamma^{k+1}, x - x^{k+1} \rangle + h(x^{k+1}) - h(x^k) - \\ &\quad - \langle \gamma^k, x^{k+1} - x^k \rangle - h(x) - h(x^k) + \langle \gamma^k, x - x^k \rangle \\ &= \langle \gamma^k - \gamma^{k+1}, x - x^{k+1} \rangle. \qquad (3.32) \end{aligned}$$

Luego, usando (3.30), se tiene:

$$D_h^{k+1}(x, x^{k+1}) + D_h^k(x^{k+1}, x^k) - D_h^k(x, x^k) \le c_k[f_X(x) - f_X(x^{k+1})] + c_k \epsilon_k. \qquad (3.33)$$

ii) Si $f_X(x) \le f_X(x^{k+1})$ usando (3.33) se deduce

$$D_h^{k+1}(x, x^{k+1}) + D_h^k(x^{k+1}, x^k) - D_h^k(x, x^k) \le c_k \epsilon_k$$

, y así

$$D_h^{k+1}(x, x^{k+1}) \le D_h^k(x, x^k) - D_h^k(x^{k+1}, x^k) + c_k \epsilon_k.$$

iii) Por (3.33) se tiene:

$$\begin{aligned} \sum_{k=1}^{l} [D_h^{k+1}(x, x^{k+1}) + D_h^k(x^{k+1}, x^k) - D_h^k(x, x^k)] &\le \sum_{k=1}^{l} c_k f_X(x) + \sum_{k=1}^{l} c_k \epsilon_k - \\ &\quad - \sum_{k=1}^{l} c_k f_X(x^{k+1}). \end{aligned}$$

$$D_h^{l+1}(x,x^{l+1}) - D_h^1(x,x^1) + \sum_{k=1}^{l} D_h^k(x^{k+1},x^k) \leq s_l f_X(x) - \sum_{k=1}^{l} c_k f_X(x^{k+1}) + \sum_{k=1}^{l} c_k \epsilon_k$$

Como $f_X(x^{l+1}) \leq f_X(x^k)$ para $k = 1, \ldots, l$ se obtiene:

$$D_h^{l+1}(x,x^{l+1}) - D_h^1(x,x^1) + \sum_{k=1}^{l} D_h^k(x^{k+1},x^k) \leq s_l[f_X(x) - f_X(x^{l+1})] + \sum_{k=1}^{l} c_k \epsilon_k.$$

$$\begin{aligned} s_l[f_X(x^{l+1}) - f_X(x)] \ &\leq \ D_h^1(x,x^1) - D_h^{l+1}(x,x^{l+1}) - \sum_{k=1}^{l} D_h^k(x^{k+1},x^k) + \\ &+ \sum_{k=1}^{l} c_k \epsilon_k. \end{aligned} \tag{3.34}$$

iv) Para obtener la expresión (3.29) dividimos ambos miembros de (3.34) por $s_l > 0$ y usamos $D_h^k(\cdot, x^k) \geq 0$.

∎

El siguiente lema establece criterios para que la sucesión $\{x^k\}$ asociada al BPM sea acotada, cada punto límite de ella pertenezca al dominio de la función h y converja a algún punto que pertenezca al dominio común de f_X y h, además la función f_X evaluada en este punto de convergencia es menor que la función f_X evaluada en cualquier punto de la sucesión.

Lema 3.2.4 *(Ver [9], Lema 4.2)*
Si $\sum_{k=1}^{\infty} c_k \epsilon_k < \infty$ *y* $x \in D_h$ *es tal que* $f_X(x^k) \geq f_X(x)$, $\forall k$, *entonces:*

i) $\{x^k\}$ *es acotada y* $\{x^k\} \subset \mathcal{L}_h^b(x,\alpha)$, *donde* $\alpha = D_h^1(x,x^1) + \sum_{k=1}^{\infty} c_k \epsilon_k$.

ii) Cada punto límite de $\{x^k\}$ *está en* $\mathcal{D}_h$.

iii) $\{x^k\}$ *converge algún* $x^\infty \in \mathcal{D}_{f_X} \cap D_h$ *tal que* $f_X(x^k) \geq f_X(x^\infty)$, $\forall k$.

Demostración .

i) Usando (3.27) se deduce:

$$D_h^l(x, x^l) \leq D_h^1(x, x^1) + \sum_{k=1}^{l-1} c_k \epsilon_k \leq \alpha, \quad \forall l.$$

Además por (3.16) $\{x^k\} \subset \mathcal{D}_{\partial h}$, por lo cual $\{x^k\} \subset \mathcal{L}_h^b(x, \alpha)$, el cual es un conjunto acotado, ya que h es de Bregman y así $\{x^k\}$ es acotada.

ii) Por la parte (i) $\{x^k\} \subset \mathcal{D}_{\partial h}$ y usando el Lema 3.1.4 se tiene que cada punto límite de $\{x^k\}$ está en $\mathcal{D}_h$.

iii) $\{x^k\}$ es acotada por la parte (i) y por tanto tiene una subsucesión $\{x^{l_j}\}$ convergente a algún x^∞, por la parte (ii) $x^\infty \in \mathcal{D}_h$.
Supongamos que $x^\infty \neq x$. Como $\{x^k\} \subset \mathcal{L}_h^b(x, \alpha)$ por la definición 3.1.1(iv) se obtiene $D_h^\sharp(x^\infty, x^{l_j}) \to 0$, entonces por (3.8) $D_h^{l_j}(x^\infty, x^{l_j}) \to 0$.
Por otro lado por (3.22) $f_X(x^{k+1}) \geq f_X(x^k)$, además f_X es cerrada y $x^{l_j} \to x^\infty$, por lo cual se puede deducir: $f_X(x^k) \geq f_X(x^\infty)$, $\forall k$.
Consideremos $l > l_j$, por (3.27) obtenemos:

$$D_h^l(x^\infty, x^l) \leq D_h^{l_j}(x^\infty, x^{l_j}) + \sum_{k=l_j}^{l-1} c_k \epsilon_k.$$

Como $\sum_{k=l_j}^{\infty} c_k \epsilon_k \to 0$ cuando $j \to \infty$ se deduce que: $D_h^l(x^\infty, x^l) \to 0$ cuando $l \to \infty$. Y por (3.8) $D_h^{'}(x^\infty, x^k) \to 0$ y usando el Lema 3.1.4(ii) se obtiene $x^k \to x^\infty$.
Si $x = x^\infty$, pero $\{x^k\}$ no es convergente. Por las partes (i) y (ii) existe un punto límite $x^{'} \in \mathcal{D}_h \backslash \{x\}$, y reemplazando x^∞ por $x^{'}$, a través del procedimiento anterior obtenemos que $x^k \to x^\infty$, lo cual es contradictorio.

∎

El siguiente teorema presenta los principales resultados de convergencia de la sucesión asociada al BPM de descenso, en el se establecen condiciones para que la sucesión $\{f_X(x^k)\}$ converja al ínfimo de f_X en el Dominio de h.

Teorema 3.2.5 *(Ver [9], Teorema 4.3)*
Supongamos que en las Hipótesis 3.2.1 (i), (ii), (iv) y (v) son válidas y h es cerrada, propia y convexa.

i) Si $\lim_{l\to\infty} \sum_{k=1}^{l} c_k\epsilon_k / \sum_{k=1}^{l} c_k = 0$, *entonces*

$$f_X(x^k) \downarrow \inf_{\mathcal{D}_h} f_X = \inf_{cl\ (\mathcal{D}_h \cap \mathcal{D}_{f_X})} f,$$

por consiguiente $f_X(x^k) \downarrow \inf_X f$ *si* $\mathcal{D}_{f_X} \subset \mathcal{D}_h$.
Si $ri\ \mathcal{D}_h \cap ri\ \mathcal{D}_{f_X} \neq \emptyset$, *es decir,* $\overset{\circ}{\mathcal{D}_h} \cap \mathcal{D}_{f_X} \neq \emptyset$, *entonces*

$$\inf_{\mathcal{D}_h} f_X = \inf_{(cl\ \mathcal{D}_h)\cap(cl\ \mathcal{D}_{f_X})} f = \inf_{cl\ \mathcal{D}_h} f_X.$$

Si $ri\ \mathcal{D}_{f_X} \subset cl\ \mathcal{D}_h$, *es decir,* $\mathcal{D}_{\partial f_X} \subset cl\ \mathcal{D}_h$, *entonces* $cl\ \mathcal{D}_{f_X} \subset cl\ \mathcal{D}_h$ *y* $Argmin_X f \subset cl\ \mathcal{D}_h$.

ii) Si h *es una B-función,* $f_X(x^k) \to \inf_{\mathcal{D}_h} f_X$, $\sum_{k=1}^{\infty} c_k\epsilon_k < \infty$ *y* $X_* = Argmin_{\mathcal{D}_h} f$ *es no vacío, entonces* $\{x^k\}$ *converge a algún* $x^\infty \in X_*$ *y* $x^\infty \in Argmin_X f$ *si* $\mathcal{D}_{f_X} \subset \mathcal{D}_h$.

iii) Si $f_X(x^k) \to \inf_{\mathcal{D}_h} f_X$, $\mathcal{D}_{f_X} \subset \mathcal{D}_h$ *y* $X_* = \emptyset$, *entonces* $|x^k| \to \infty$.

Demostración .

i) Sea $x \in \mathcal{D}_h$, si tomamos límite en 3.29 obtenemos

$$\lim_{l\to\infty} f_X(x^l) \leq f_X(x),$$

ya que por (3.22) $f_X(x^{l+1}) \leq f_X(x^l)$, por Hipótesis 3.2.1 (v), $s_l \to \infty$ y por hipótesis $\sum_{k=1}^{l} c_k\epsilon_k/s_l \to 0$.
Luego, usando ([16], Corolario 7.3.2), se deduce:

$$f_X(x^k) \to \inf_{\mathcal{D}_h} f_X = \inf_{\mathcal{D}_h \cap \mathcal{D}_{f_X}} f = \inf_{cl(\mathcal{D}_h \cap \mathcal{D}_{f_X})} f.$$

Si $\mathrm{ri}\mathcal{D}_h \cap \mathrm{ri}\mathcal{D}_{f_X} \neq \emptyset$, entonces $\mathrm{cl}(\mathcal{D}_h \cap \mathcal{D}_{f_X}) = \mathrm{cl}(\mathcal{D}_h) \cap \mathrm{cl}(\mathcal{D}_{f_X})$ ([16], Teorema 6.5)
Por lo cual

$$\inf_{\mathcal{D}_h} f_X = \inf_{\mathrm{cl}(\mathcal{D}_h)\cap \mathrm{cl}(\mathcal{D}_{f_X})} f \leq \inf_{\mathcal{D}_{f_X} \cap \mathrm{cl}(\mathcal{D}_h)} f = \inf_{\mathrm{cl}(\mathcal{D}_h)} f_X$$

Y así $\inf_{\mathcal{D}_h} f_X = \inf_{\text{cl}(\mathcal{D}_h)} f_X$.

Si $\text{ri}\mathcal{D}_{f_X} \subset \text{cl}\mathcal{D}_h$, usando ([16], Teorema 6.5) se tiene $\text{cl}\mathcal{D}_{f_X} \subset \text{cl}\mathcal{D}_h$.

ii) Si $x \in X_*$, entonces $f_X(x^k) \to f_X(x)$, pero $f_X(x^k) \geq f_X(x)$, $\forall k$, ya que por (3.16), $x^k \in \mathcal{D}_h \cap \mathcal{D}_{f_X}$, de donde se puede deducir por el Lema 3.2.4 que: $x^k \to x^\infty \in \mathcal{D}_h \cap \mathcal{D}_{f_X}$ y $\lim_{k\to\infty} f_X(x^k) \geq f_X(x)$, por lo cual $x^\infty \in X_*$.

iii) Si $|x^k| \nrightarrow \infty$, $\{x^k\}$ tiene un punto límite x con $f_X(x) \leq \inf_{\mathcal{D}_h} f_X$, pero $f_X(x^k) \to \inf_{\mathcal{D}_h} f_X$ y como $\mathcal{D}_{f_X} \subset \mathcal{D}_h$ se tiene $x \in \mathcal{D}_h \cap X_*$, por lo tanto $X_* \neq \emptyset$, lo cual es contradictorio, queda demostrado que $|x^k| \to \infty$.

∎

Lema 3.2.6 *(Ver [9], Lema 4.5)*

i) Si $\epsilon_k \to 0$, *entonces* $\sum_{k=1}^{l} c_k\epsilon_k/s_l \to 0$, *cuando* $l \to \infty$.

ii) Si $\sum_{k=1}^{\infty} \epsilon_k < \infty$ *y* $\{c_k\} \subset (0, c_{\max}]$, *para algún* $c_{\max} < \infty$, *entonces*

$$\sum_{k=1}^{\infty} c_k\epsilon_k < \infty.$$

Demostración .

i) Sea $\epsilon > 0$, elegimos un $\bar{k}$ tal que $\epsilon_k \leq \epsilon$, $\forall k \geq \bar{k}$.
Como $s_l \to \infty$, podemos seleccionar un $\bar{l} > \bar{k}$ tal que:

$$\sum_{k=1}^{\bar{k}} c_k\epsilon_k/s_l \leq \epsilon, \quad \forall l \geq \bar{l},$$

entonces:

$$\sum_{k=1}^{l} c_k\epsilon_k/s_l \leq \sum_{k=1}^{\bar{k}} c_k\epsilon_k/s_l + \epsilon \sum_{k=\bar{k}+1}^{l} c_k/\sum_{k=1}^{l} c_k \leq 2\epsilon \quad \forall l \geq \bar{l}.$$

ii) Tenemos que $\sum_{k=1}^{\infty} c_k\epsilon_k \leq c_{\max} \sum_{k=1}^{\infty} \epsilon_k < \infty$.

∎

3-3 MÉTODO DE MULTIPLICADORES INEXACTOS

Asociado al BPM Kiwiel presenta en [9] el método de multiplicadores inexactos, el cual resuelve el problema dual del problema resuelto por el BMP. Un método similar propone Castillo en [3], llamado método de multiplicadores basados en shifts. En el capítulo siguiente se aplicará esté método propuesto por Castillo, usando una penalidad no coerciva especifica.

El problema que resuelve este método de multiplicadores inexactos es el siguiente:

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeto a} & g_i(x) \leq 0, \quad i = 1, \ldots, m. \end{array} \tag{3.35}$$

Donde $f, g_1, g_2, \ldots, g_m$ son cerradas, propias y convexas en $\mathbb{R}^n$ con:

$$\mathcal{D}_f \subset \cap_{i=1}^m \mathcal{D}_{g_i} \quad \text{y} \quad \text{ri } \mathcal{D}_f \subset \cap_{i=1}^m \text{ri} \mathcal{D}_{g_i}. \tag{3.36}$$

Siendo $g(\cdot) = (g_1(\cdot), \ldots, g_m(\cdot))$.

Definimos el Lagangiano de (3.35) de la siguiente manera:

$$L(x, \pi) = \begin{cases} f(x) + \langle \pi, g(x) \rangle & \text{si } \; x \in \mathcal{D}_f \quad \text{y} \quad \pi \in \mathbb{R}^m_+ \\ -\infty & \text{si } \; x \in \mathcal{D}_f \quad \text{y} \quad \pi \notin \mathbb{R}^m_+ \\ \infty & \text{si } \; x \notin \mathcal{D}_f. \end{cases} \tag{3.37}$$

Podemos definir la función dual $d(\pi) = \inf_x L(x, \pi)$. Entonces $d(\pi) = -\infty$ si $\pi \notin \mathbb{R}^m_+$. Asumamos que $d(\pi) > -\infty$, para algún π. El problema dual de (3.35) es maximizar d o equivalentemente minimizar $q(\pi)$ con $\pi \geq 0$, donde $q = -d$ es una función cerrada propia y convexa.

Tomemos una B-función h en $\mathbb{R}^m$, asumamos que $\mathbb{R}^m_> \subset \mathcal{D}_h$, consideremos la función:

$$h_+ = h + \iota_{\mathbb{R}^m_>},$$

la cual es una B-función (Ver Lema 3.1.1). Debemos notar que nuestra penalidad definida en (3.15) no verifica esta condición, en virtud de lo cual no está incluida en el análisis realizado por Kiwiel en [9].

La penalidad que usa Kiwiel en su método de multiplicadores inexactos es la monótona conjugada de h, la cual definimos de la siguiente manera:

$$h^+(\cdot) = \sup_{\pi \geq 0}\{\langle \pi, \cdot\rangle - h(\pi)\}.$$

Esta función es no decreciente, ya que sí $u \leq u'$, entonces $\langle \pi, u\rangle \leq \langle \pi, u'\rangle, \forall \pi \geq 0$, de donde

$$h^+(u) = \sup_{\pi \geq 0}\{\langle \pi, u\rangle - h(\pi)\} \leq \sup_{\pi \geq 0}\{\langle \pi, u'\rangle - h(\pi)\} = h^+(u').$$

Además coincide con la conjugada h_+^* de h_+, ya que

$$h^+(\cdot) = \sup_{\pi}\{\langle \pi, \cdot\rangle - h_+(\pi)\} = h_+^*(\cdot).$$

Antes de presentar el método de multiplicadores inexactos consideramos el siguiente lema, el cual establece características importantes de las funciones h^+ y h_+, estas características son derivadas de las que posee la función h, la cual es una B-función y en consecuencia h_+ también lo es y la monótona conjugada de h es esencialmente suave.

Lema 3.3.1 *(Ver [9], Lema 7.2, Apéndice A)*
Si h es una función cerrada, propia y esencialmente estrictamente convexa en $\mathbb{R}^m$ con $\mathbb{R}^m_+ \cap ri\mathcal{D}_h \neq \emptyset$, entonces h^+ es cerrada, propia convexa y esencialmente suave, $\partial h^+(u) = \{\nabla h^+(u)\}$ para cada $u \in \mathcal{D}_{\partial h^+}$, $\partial h^+ = (\partial h_+)^{-1}$ y ∇h^+ es continua en $\mathcal{D}_{\partial h^+} = \overset{\circ}{\mathcal{D}}_{h^+} = im\partial h_+$.
Además $\mathcal{D}_{h^+} = \mathcal{D}_{h^+} - \mathbb{R}^m_+$, $\overset{\circ}{\mathcal{D}}_{h^+} = \overset{\circ}{\mathcal{D}}_{h^+} - \mathbb{R}^m_+$, $\partial h_+ = \partial h + N_{\mathbb{R}^m_+}$ y $\nabla h^+ = \nabla h^+ \circ (I + N_{\mathbb{R}^m_+} \circ \nabla h^+)$, donde I es el operador identidad y $N_{\mathbb{R}^m_+} = \partial \iota_{\mathbb{R}^m_+}$ es

el operador cono normal de $\mathbb{R}^m_+$*, es decir* $N_{\mathbb{R}^m_+}(\pi) = \{\gamma \leq 0 \ : \ \langle \gamma, \pi \rangle = 0\}$*, si* $\pi \geq 0$*,* $N_{\mathbb{R}^m_+}(\pi) = \emptyset$*, si* $\pi \not\geq 0$*. Si adicionalmente* $im\partial h \supset \mathbb{R}^m_>$*, entonces* h_+ *es cofinita,* $\mathcal{D}_{h^+} = \mathbb{R}^m$*, y* h^+ *es continuamente diferenciable.*

3-3-1 MÉTODO DE MULTIPLICADORES INEXACTOS:

Como $\mathbb{R}^m_> \subset \mathcal{D}_h$, podemos encontrar $\inf_{\pi \geq 0} q(\pi)$ a través del método BPM, reemplazando en (3.16)-(3.21) f, X, h y x^k por $q, \mathbb{R}^m, h_+$ y π^k respectivamente.

El método de multiplicadores inexactos requiere encontrar dos sucesiones $\{x^k\}$ y $\{\pi^k\}$, procediendo de la siguiente manera:

Dados $\pi^k \in \mathcal{D}_q \cap \mathcal{D}_{\partial h_+}$ y $\gamma^k \in \partial h_+(\pi^k)$.

Encontrar π^{k+1} y x^{k+1} tal que:

$$L(x^{k+1}, \pi^{k+1}) \leq \inf_x L(x, \pi^{k+1}) + \epsilon_k = d(\pi^{k+1} + \epsilon_k), \tag{3.38}$$

$$\pi^{k+1} = \nabla h^+(\gamma^k + c_k g(x^{k+1})). \tag{3.39}$$

$$p^{k+1} \in \partial_{\epsilon_k} q(\pi^{k+1}), \tag{3.40}$$

$$\gamma^{k+1} = \gamma^k - c_k p^{k+1} \in \partial_{h_+}(\pi^{k+1}). \tag{3.41}$$

Para algún p^{k+1} y γ^{k+1}.

Notemos que (3.38) implica:

$$-g(x^{k+1}) \in \partial_{\epsilon_k} q(\pi^{k+1}) = \partial_{\epsilon_k} q(\pi^{k+1}) + \partial \iota_{\mathbb{R}^m_+}(\pi^{k+1}). \tag{3.42}$$

En efecto: si tomamos un π arbitrario tenemos:

$$q(\pi) = \sup_x \{-L(x, \pi)\} \geq -L(x^{k+1}, \pi) = -f(x^{k+1}) - \langle \pi, g(x^{k+1}) \rangle.$$

Además por (3.38) se deduce:

$$\begin{aligned} q(\pi) &\geq -f(x^{k+1}) - \langle \pi, g(x^{k+1}) \rangle \\ &= -f(x^{k+1}) - \langle \pi^{k+1}, g(x^{k+1}) \rangle + \langle -g(x^{k+1}), \pi - \pi^{k+1} \rangle \\ &= -L(x^{k+1}, \pi^{k+1}) + \langle -g(x^{k+1}), \pi - \pi^{k+1} \rangle \\ &\geq q(\pi^{k+1}) - \epsilon_k + \langle -g(x^{k+1}), \pi - \pi^{k+1} \rangle. \end{aligned}$$

Por otro lado, como $\mathcal{D}_q \subset \mathbb{R}^m_+$, se deduce (3.42).
A partir de (3.39), deducimos que $\pi^{k+1} \in \mathcal{D}_{\partial h_+} \subset \mathcal{D}_h \subset \mathbb{R}^m_+$, mientras

$$q(\pi^{k+1}) \geq q(\pi^k) + \epsilon_k,$$

ya que $\pi^{k+1} \in \mathcal{D}_q$.

Por (3.40); (3.41) y (3.42) tomando $p^{k+1} = (\gamma^k - \gamma^{k+1})/c_k$ se deduce:

$$\gamma^{k+1} = \gamma^k + c_k g(x^{k+1}) - \tilde{\gamma}^{k+1} \in \partial h_+(\pi^{k+1}) \quad \text{con} \quad \tilde{\gamma}^{k+1} \in N_{\mathbb{R}^m_+}(\pi^{k+1}). \quad (3.43)$$

Entonces de (3.41) y (3.43) se infiere:

$$p^{k+1} = -g(x^{k+1}) + \tilde{\gamma}^{k+1}/c_k \in \partial_{\epsilon_k} q(\pi^{k+1}). \quad (3.44)$$

Usando (3.39) y $(\partial h_+)^{-1} = \nabla h^+$ (ver Lema 3.3.1) se tiene:

$$\gamma^k + c_k g(x^{k+1}) \in \partial h_+(\pi^{k+1}) = \partial h(\pi^{k+1}) + N_{\mathbb{R}^m_+}(\pi^{k+1}). \quad (3.45)$$

Usaremos $\tilde{\gamma}^{k+1} = 0$, por lo que obtendremos $p^{k+1} = -g(x^{k+1})$.
Para facilitar la notación escribimos (3.39) de la siguiente manera:

$$\pi^{k+1} = \nabla P_k(g(x^{k+1})), \quad (3.46)$$

donde

$$P_k(u) = h^+(\gamma^k + c_k u)/c_k, \quad \forall u \in \mathbb{R}^m. \quad (3.47)$$

Sea

$$L_k(x) = f(x) + \frac{1}{c_k}[h^+(\gamma^k + c_k g(x)) - h^+(\gamma^k)]. \quad (3.48)$$

Si $x \in \mathcal{D}_f \subset \mathcal{D}_g = \cap_{i=1}^m \mathcal{D}_{g_i}$ y $L_k(x) = \infty$ de otra forma.

El siguiente lema establece las condiciones para que la función L_k sea propia y convexa, de igual modo modo se presenta la manera de calcular el subdiferencial de L_k, más aun se prueba que este subdiferencial coincide con el subdiferencial, respecto de x, de la función Lagrangiano.

Lema 3.3.2 *(Ver [9], Lema 7.3)*
Supongamos que $\inf_{\mathcal{D}_f} \max_{i=1}^m g_i \geq 0$, *el conjunto factible del problema (3.35),* $C_0 = \{x \in \mathcal{D}_f : g(x) \leq 0\}$, *es no vacío. Entonces la función* L_k *es propia y convexa y se cumple:*

$$\partial L_k(x) = \partial f(x) + \sum_{i=1}^{m} [\nabla P_k(g(x))]_i \partial g_i(x) \quad \forall x \in \mathcal{D}_{L_k} \supset \mathcal{D}_{\partial L_k}. \tag{3.49}$$

Sí $\partial L_k(x) \neq \emptyset$, *entonces* $\pi = \nabla P_k(g(x))$ *está bien definido,* $\pi \geq 0$ *y* $\partial L_k(x) = \partial_x L(x, \pi)$, *donde*

$$\partial_x L(x, \pi) = \partial f(x) + \sum_{i=1}^{m} \pi_i \partial g_i(x), \quad \forall x \in \mathbb{R}^n, \forall \pi \in \mathbb{R}^m_+. \tag{3.50}$$

Sí $\hat{x} \in Argmin L_k$, *entonces* $\hat{x} \in Argmin_x L(x, \hat{\pi})$, *para* $\hat{\pi} = \nabla P_k(g(\hat{x}))$.
La aseveración anterior es válida cuando $\inf_{\mathcal{D}_f} \max_{i=1}^m g_i > 0$, *pero* $\mathcal{D}_{h^+} = \mathbb{R}^m$. *Por ejemplo si* $im\partial h \supset \mathbb{R}^m_>$. *(ver Lema 3.3.1)*

Demostración . Usando $\gamma^k \in \partial h_+(\pi^k) \subset \overset{\circ}{\mathcal{D}}_{h^+}$, (ver Lema 3.3.1), además como $\overset{\circ}{\mathcal{D}}_{P_k} = (\overset{\circ}{\mathcal{D}}_{h^+} - \gamma^k)/c_k$, escogemos $\tilde{u} \in \mathcal{D}_{P_k} \cap \mathbb{R}^m_>$ y $\tilde{x} \in \mathcal{D}_f$, tal que $g(\tilde{x}) < \tilde{u}$. Entonces como P_k es no decreciente, dado que h^+ es nodecreciente, por otro lado de (3.36) se tiene $\mathrm{ri}\mathcal{D}_f \subset \cap_i \mathrm{ri}\mathcal{D}_{g_i}$. Por ([9] Lema A.1) obtenemos $\mathrm{im}\partial P_k \subset \mathbb{R}^m_+$ y la ecuación 3.49 se obtiene usando $\partial P_k = \{\nabla P_k\}$ (Lema 3.3.1).
Por consiguiente sí $\partial L_k(x) \neq \emptyset$, entonces $\pi = \nabla P_k(g(x)) \geq 0$, así $\mathrm{ri}\mathcal{D}_f \subset \cap_i \mathrm{ri}\mathcal{D}_{g_i}$ implica:

$$\partial_x L(x, \pi) = \partial f(x) + \sum_{i=1}^{m} \pi_i \partial g_i(x) = \partial L_k(x).$$

Sí $\hat{x} \in \mathrm{Argmin} L_k$, entonces $0 \in \partial L_k(\hat{x}) = \partial_x L(\hat{x}, \hat{\pi})$, para $\hat{\pi} = \nabla P_k(g(\hat{x}))$, por lo cual $\hat{x} \in \mathrm{Argmin}_x L(x, \hat{\pi})$.
Finalmente cuando $\mathcal{D}_{h^+} = \mathbb{R}^m$, para cualquier $\tilde{x} \in \mathcal{D}_f$ se puede escoger un $\tilde{u} \in \mathcal{D}_{P_k}$ con $g(\tilde{x}) < \tilde{u}$, ya que $\mathcal{D}_f \subset \cap_i \mathcal{D}_{g_i}$ y $\mathcal{D}_{P_k} = \mathbb{R}^m$.

∎

Consideremos ahora el algoritmo asociado al método de multiplicadores inexactos presentados por Kiwiel en [9].

Algoritmo del método de multiplicadores inexactos (ver Algoritmo 7.4, [9])

Para la iteración $k \geq 1$, dados $\pi^k \in \mathcal{D}_q$ y $\gamma^k \in \partial h_+(\pi^k)$,
encontrar:

$$x^{k+1} \in \mathrm{Argmin}_{x \in \mathcal{D}_f} \left\{ f(x) + \frac{1}{c_k} h^+(\gamma^k + c_k g(x)) \right\}.$$

$$\pi^{k+1} = \nabla h^+(\gamma^k + c_k g(x^{k+1})).$$

Tal que se cumpla (3.38),
elegimos γ^{k+1} que satisfaga (3.43) y $p^{k+1} = (\gamma^k - \gamma^{k+1})/c_k$.

Un método similar a este desarrollaremos en el siguiente capítulo, no obstante usaremos una función de penalidad no coerciva, la cual denominaremos θ, en lugar de usar la monótona conjugada de h.

3-3-2 CONVERGENCIA DEL MÉTODO DE MULTIPLICADORES INEXACTOS

En este método tenemos dos sucesiones $\{x^k\}$ y $\{\pi^k\}$ las cuales se generan por el algoritmo de multiplicadores inexactos, la convergencia de la sucesión $\{\pi^k\}$ ya ha sido estudiada en la sección 3.2.2, ya que esta sucesíon coincide con la sucesión generada por el BPM y con respecto a la convergencia de $\{x^k\}$ introduciremos el concepto de convergencia ergódica. Asumamos que son válidas las siguientes hipótesis:

Hipótesis 3.3.1

i) h_+ *es una B-función tal que* $\mathcal{D}_h \supset \mathbb{R}^m_>$.

ii) $\mathcal{D}_q \cap \mathbb{R}^m_> \neq \emptyset$ *ó* $\emptyset \neq \mathcal{D}_q \subset \mathcal{D}_{h_+}$, *donde* $-q = d = \inf_x L(x, \cdot)$.

iii) $\{c_k\}$ *es una sucesión de números positivos tales que* $s_k = \sum_{j=1}^{k} c_j \to \infty$.

Al considerar estas hipótesis tenemos que q es cerrada, propia y convexa. Además

$$\overset{\circ}{\mathcal{D}}_{h_+} = \mathbb{R}^m_> \subset \mathcal{D}_{h_+} \subset \mathbb{R}^m_+.$$

Por lo cual

$$\mathrm{cl}\mathcal{D}_{h_+} = \mathbb{R}^m_+ \supset \mathcal{D}_q.$$

Por otro lado $\mathcal{D}_q \cap \overset{\circ}{\mathcal{D}}_{h_+} \neq \emptyset$ sí $\mathcal{D}_q \cap \mathbb{R}^m_> \neq \emptyset$ y $\inf_{\mathcal{D}_{h_+}} q = \inf q = \inf_{\mathrm{cl}\mathcal{D}_{h_+}} q$. Por consiguiente podemos aplicar el método BMP al problema dual $\sup d = -\inf q$ con la B-función h_+. De esta manera usamos los resultados relacionados con este método obtenidos en las secciones 3.2.1 y 3.2.2, para ello se reemplaza a f, X y h por $q, \mathbb{R}^m$ y h_+.

En el siguiente teorema se establecen condiciones para que la sucesión $\{\pi^k\}$ generada por el algoritmo del método de multiplicadores inexactos converja.

Teorema 3.3.3 *(Ver [9], Teorema 8.3)*
Sí $\sum_{j=1}^{k} s_j\epsilon_j/s_k \to 0$*, entonces* $d(\pi^k) \to \sup d$*. Sí* $d(\pi^k) \to \sup d$*,* $\mathcal{D}_{h_+} \cap$ *Argmax* $d \neq \emptyset$ *y* $\sum_{k=1}^{\infty} c_k\epsilon_k < \infty$*, entonces* $\pi^k \to \pi^\infty \in$ *Argmax* d*. Sí* $d(\pi^k) \to \sup d$*,* $\mathcal{D}_q \subset \mathcal{D}_{h_+}$ *y* *Argmax*$_{\mathcal{D}_{h_+}}$ $d = \emptyset$ *es decir* $\mathcal{D}_{h_+} = \mathbb{R}^m_+$ *y* *Argmax*$_{\mathcal{D}_{h_+}} d = \emptyset$ *entonces* $|\pi^k| \to \infty$*.*

El siguiente teorema establece condiciones para que $d(\pi^k) \to \sup d$, así como para establecer que $\limsup_{k\to\infty} f(x^k) \leq \sup_{\pi} d(\pi)$ *y* $\limsup_{k\to\infty} g_i(x^k) \leq 0$.

Teorema 3.3.4 *(Ver [9], teorema 8.4)*
Sea $\mathcal{D}_{\nabla h} \supset \mathbb{R}^m_+$*,* $\gamma^k = \nabla h(\pi^k)$ $\forall k$ *y* $\sum_{j=1}^{k} s_j\epsilon_j/s_k \to 0$*, entonces* $d(\pi^k) \to \sup d$*. Si Argmax* $d \neq \emptyset$ *y* $\sum_{k=1}^{\infty} c_k\epsilon_k < \infty$*, entonces* $\pi^k \to \pi^\infty \in$ *Argmax* d*. Si* $\inf_k c_k > 0$*, entonces*

$$\limsup_{k\to\infty} f(x^k) \leq \sup_{\pi} d(\pi) \quad y \quad \limsup_{k\to\infty} g_i(x^k) \leq 0, \quad i = 1, \dots, m. \tag{3.51}$$

Además cada punto límite de $\{x^k\}$ *resuelve a 3.35. Si Argmax* $d \neq \emptyset$*, entonces* $|\pi^k| \to \infty$*.*

Demostración . Como $\sum_{j=1}^{k} s_j \epsilon_j / s_k \to 0$, por el Teorema 3.3.3 se deduce $d(\pi^k) \to \sup d$. Por otro lado $\mathcal{D}_h \supset \mathcal{D}_{\nabla h} \supset \mathbb{R}^m_+$, de donde $\mathcal{D}_{h_+} = \mathbb{R}^m_+$. Además $\emptyset \neq$ Argmin $d \subset \mathbb{R}^m_+$, por lo cual $\mathcal{D}_{h_+} \bigcap$ Argmin $d \neq \emptyset$, usamos nuevamente el teorema 3.3.3 para obtener $\pi^k \to \pi^\infty \in$ Argmax d.
Supongamos que $\pi^k \to \pi^\infty \in$ Argmax d, además $\inf_k c_k > 0$. Como $p^k = (\gamma^{k-1} - \gamma^k)/c_{k-1}$, con $p^k + g(x^k) \in N_{\mathbb{R}^m_+}(\pi^k)$. Entonces por el Lema 3.3.1 tenemos $\langle \pi^k, g(x^k)\rangle = -\langle \pi^k, p^k\rangle$ y $g(x^k) \leq -p^k$, $\forall k > 1$. Como $\pi^k \to \pi^\infty$, ∇h es continua en $\mathbb{R}^m_+$ y $c_k \geq c_{\min}$ $\forall k$, se deduce $p^k \to 0$. Por consiguiente $\langle \pi^k, g(x^k)\rangle \to 0$ y $\limsup\limits_{k\to\infty} g_i(x^k) \leq 0$ $\forall i$.
Por otro lado de (3.38) se tiene que $L(x^k, \pi^k) \leq \inf_x L(x, \pi^k) + \epsilon_{k-1}$, de donde $f(x^k) + \langle \pi^k, g(x^k)\rangle \leq f(x) + \langle \pi^k, g(x)\rangle + \epsilon_{k-1}$ $\forall x$, y como $\langle \pi^k, g(x^k)\rangle \to 0$, $\epsilon_k \to 0$ y $\pi^k \to \pi^\infty$ al tomar límite se obtiene $\limsup_k f(x^k) \leq L(x, \pi^\infty)\ \forall x$. Y así $\limsup_k f(x^k) \leq d(\pi^\infty)$.
Supongamos que $x^k \xrightarrow{K} x^\infty$ para algún x^∞ y $K = \{1, 2, \ldots\}$, por (3.51) tenemos $f(x^\infty) \leq \sup d$ y $g(x^\infty) \leq 0$, ya que f y g son cerradas. Por otro lado $d(\pi) \leq f(x^\infty) + \langle \pi, g(x^\infty)\rangle \leq f(x^\infty)$ $\forall \pi \in \mathbb{R}^m_+$. Entonces $f(x^\infty) \geq \sup d$ por lo tanto $f(x^\infty) = \max d$, y así x^∞ resuelve el problema (3.35).

$\blacksquare$

El siguiente lema establece la siguiente relación de desigualdad $L(x^{k+1}, \pi^k) \leq L_k(x^{k+1}) \leq L(x^{k+1}, \pi^{k+1})$, entre la función lagrangiana L y la función L_k, además se demuestra que $d(\pi^k) \leq d(\pi^{k+1}) + \epsilon_k$.

Lema 3.3.5 *(Ver [9], Lema 8.9)*
Sí $u^{k+1} = g(x^{k+1})$ $\forall k$, *tenemos:*

$$L(x^{k+1}, \pi^{k+1}) = L_k(x^{k+1}) + D_{h^+}(\gamma^k, \gamma^k + c_k u^{k+1})/c_k \geq L_k(x^{k+1}), \quad (3.52)$$

$$L_k(x^{k+1}) = L(x^{k+1}, \pi^k) + D_{h^+}(\gamma^k + c_k u^{k+1}, \gamma^k)/c_k \geq L(x^{k+1}, \pi^k), \quad (3.53)$$

$$\begin{aligned} L(x^{k+1}, \pi^{k+1}) - L(x^{k+1}, \pi^k) &= \langle \pi^{k+1} - \pi^k, u^{k+1}\rangle \\ &= \langle \nabla h^+(\gamma^k + c_k u^{k+1}) - \nabla h^+(\gamma^k), u^{k+1}\rangle \\ &\geq 0. \end{aligned} \quad (3.54)$$

$$d(\pi^k) \leq L(x^{k+1}, \pi^k) \leq L_k(x^{k+1}) \leq L(x^{k+1}, \pi^{k+1}) \leq d(\pi^{k+1}) + \epsilon_k. \quad (3.55)$$

Demostración . Probemos (3.52), usando (3.48), (3.39), (3.37) y (2.36) se obtiene:

$$\begin{aligned}
L(x^{k+1}, \pi^{k+1}) &= f(x^{k+1}) + \langle \pi^{k+1}, u^{k+1} \rangle \\
&= L_k(x^{k+1}) - \frac{1}{c_k}[h^+(\gamma^k + c_k u^{k+1}) - h^+(\gamma^k)] + \langle \pi^{k+1}, u^{k+1} \rangle \\
&= L_k(x^{k+1}) + \frac{1}{c_k}[h^+(\gamma^k) - h^+(\gamma^k + c_k u^{k+1}) + \langle \pi^{k+1}, c_k u^{k+1} \rangle] \\
&= L_k(x^{k+1}) + D_{h^+}(\gamma^k, \gamma^k + c_k u^{k+1})/c_k \\
&\geq L_k(x^{k+1}).
\end{aligned}$$

Para probar (3.53) usamos el Lema 3.3.1 y obtenemos $\nabla h^+ = (\partial h_+)^{-1}$ y como $\gamma^k \in \partial h_+(\pi^k)$, se deduce $\pi^k = \nabla h^+(\gamma^k)$, por lo cual:

$$\begin{aligned}
L_k(x^{k+1}) &= f(x^{k+1}) + \frac{1}{c_k}[h^+(\gamma^k + c_k u^{k+1}) - h^+(\gamma^k)] \\
&= L(x^{k+1}, \pi^k) + \frac{1}{c_k}[h^+(\gamma^k + c_k u^{k+1}) - h^+(\gamma^k) - \langle \pi^k, c_k u^{k+1} \rangle] \\
&= L(x^{k+1}, \pi^k) + \frac{1}{c_k}[h^+(\gamma^k + c_k u^{k+1}) - h^+(\gamma^k) - \langle \nabla h^+(\gamma^k), c_k u^{k+1} \rangle] \\
&= L(x^{k+1}, \pi^k) + D_{h^+}(\gamma^k + c_k u^{k+1}, \gamma^k)/c_k \\
&\geq L(x^{k+1}, \pi^k).
\end{aligned}$$

Para probar (3.54) usamos (3.37) y (3.39), obteniendose:

$$\begin{aligned}
L(x^{k+1}, \pi^{k+1}) - L(x^{k+1}, \pi^k) &= f(x^{k+1}) + \langle \pi^{k+1}, u^{k+1} \rangle - f(x^{k+1}) - \langle \pi^k, u^{k+1} \rangle \\
&= \langle \nabla h^+(\gamma^k + c_k u^{k+1}) - \nabla h^+(\gamma^k), u^{k+1} \rangle \\
&\geq 0.
\end{aligned}$$

Finalmente para probar (3.55) se usa (3.52),(3.53),(3.54) y (3.38)

$$d(\pi^k) \leq L(x^{k+1}, \pi^k) \leq L_k(x^{k+1}) \leq L(x^{k+1}, \pi^{k+1}) \leq d(\pi^{k+1}) + \epsilon_k.$$

∎

Con el fin de presentar un resultado equivalente al Teorema 3.3.4, pero sin solicitar la condición $\mathcal{D}_{\nabla h} \supset \mathbb{R}^m_+$, se introduce a continuación la estructura ergódica, para cada k se define una sucesión agregada a la solución primal:

$$\breve{x}^{k+1} = \sum_{j=1}^{k} c_j x^{j+1}/s_k \quad \text{donde} \quad s_k = \sum_{j=1}^{k} c_j. \tag{3.56}$$

Debemos notar que $\sum\limits_{j=1}^{k} c_j/s_k = 1$, por lo cual usando la Desigualdad de Jensen (Teorema 1.1.3), dado que g es convexa se obtiene

$$g(\breve{x}^{k+1}) \leq \sum_{j=1}^{k} c_j g(x^{j+1})/s_k,$$

además por (3.43) y (3.44) se deduce que $c_j g(x^{j+1}) \leq -c_j p^{j+1} = \gamma^{j+1} - \gamma^j$ para $j = 1, \ldots, k$. Por lo cual:

$$g(\breve{x}^{k+1}) \leq \sum_{j=1}^{k} c_j g(x^{j+1})/s_k \leq (\gamma^{k+1} - \gamma^1)/s_k. \tag{3.57}$$

Lema 3.3.6 *(Ver [9], Lema 8.10)*
Supongamos que $\sup_{i,k} \gamma_i^k < \infty$, $\epsilon_k \to 0$, $\langle \pi^k, u^k \rangle \to 0$ *y* $d(\pi^k) \to d^\infty < \infty$*, entonces:*

$$\limsup_{k\to\infty} f(\breve{x}^k) \leq d^\infty \text{ } y \text{ } \limsup_{k\to\infty} g_i(\breve{x}^k) \leq 0, \quad i = 1, \ldots, m. \tag{3.58}$$

Si $\{\breve{x}^k\}$ *tiene un punto límite* x^∞ *, entonces* x^∞ *resuelve el problema (3.35),* $f(x^\infty) = d^\infty = \max d$ *y cada punto límite de* $\{\pi^k\}$ *maximiza a* d.

Demostración . Como $s_k \to \infty$, por (3.57) se tiene $\limsup_k g_i(\breve{x}^k) \leq 0$, $\forall i$. Por (3.56) y usando la convexidad de f se deduce $f(\breve{x}^{k+1}) \leq \sum\limits_{i=1}^{k} c_j f(x^{j+1})/s_k$, por otro lado por (3.55) tenemos $f(x^k) = L(x^k, \pi^k) - \langle \pi^k, u^k \rangle \to d^\infty$, ya que $\langle \pi^k, u^k \rangle \to 0$ y $d(\pi^k) \to d^\infty < \infty$. Por tanto $\limsup_k f(\breve{x}^k) \leq d^\infty$.
Supongamos que $\breve{x}^k \xrightarrow{K} x^\infty$, como f y g son cerradas, usando (3.58) obtenemos $f(x^\infty) \leq d^\infty$ y $g(x^\infty) \leq 0$. Por otro lado por dualidad se tiene

$$d(\pi^k) \to d^\infty \leq f(x^\infty),$$

por lo cual $f(x^\infty) = d^\infty = \max d$, y x^∞ resuelve el problema 3.35. Además como $d(\pi^k) \to d^\infty$ y d es cerrada, entonces cada punto de acumulación de $\{\pi^k\}$ maximiza a d. ∎

3-4 FUNCIONES DE PENALIDAD

Estudiaremos ahora una clasificación de las funciones de penalidad, consideradas por Kiwiel en [9], así como sus correspondientes funciones de Bregman asociadas. Las funciones de penalidad son utilizadas en los métodos de multiplicadores, mientras las de funciones de Bregman son usadas en el Método Proximal (BPM). A continuación son presentadas algunas familias específicas de penalidad consideradas por Kiwiel en [9], de igual modo se presentan las familias de las funciones de Bregman asociadas a estas.

Definición 3.4.1 *(Ver [9], Definición 9.1)*
Decimos que $\psi \in \Psi$, si sólo si ψ es B-función en $\mathbb{R}$ con $\mathcal{D}_\psi \supset \mathbb{R}_>$.
Definimos las siguientes familias de funciones: $\tilde{\Psi} = \{\psi \in \Psi : \mathcal{D}_{\nabla\psi} \supset \mathbb{R}_>\}$, $\Psi_0 = \{\psi \in \tilde{\Psi} : \mathcal{D}_{\partial\psi} \supset \mathbb{R}_+\}$, $\Psi_s = \{\psi \in \tilde{\Psi} : \mathcal{D}_{\partial\psi} = \mathbb{R}_>\}$, $\Psi_{s'} = \{\psi \in \Psi_s : 0 \in \mathcal{D}_\psi\}$ y $\Psi_{s''} = \{\psi \in \Psi_s : 0 \notin \mathcal{D}_\psi\}$.

Debemos notar que $\tilde{\Psi} = \Psi_0 \cup \Psi_s$, ya que si $\psi \in \Psi_0 \cup \Psi_s$, obviamente se tiene que $\psi \in \tilde{\Psi}$. Por otro lado si $\psi \in \tilde{\Psi}$, se tiene $\mathcal{D}_{\partial\psi} \supset \mathcal{D}_{\nabla\psi} \supset \mathbb{R}_>$, entonces se deduce que $\psi \in \Psi_0$ ($0 \in \mathcal{D}_{\partial\psi}$) ó $\psi \in \Psi_s$ ($0 \notin \mathcal{D}_{\partial\psi}$); por consiguiente $\psi \in \Psi_0 \cup \Psi_s$.
Además $\psi_+ = \psi$, si $\psi \in \Psi_s$, ya que $\mathcal{D}_\psi \subset \mathrm{cl}\, \mathcal{D}_{\partial\psi} = \mathbb{R}_+$.

Definición 3.4.2 *(Ver [9],Definición 9.3)*
Decimos que $\phi \in \Phi$ si sólo si $\phi : \mathbb{R} \to (-\infty, \infty]$ es cerrada, propia, convexa y esencialmente suave y se cumple:

$$\mathring{\mathcal{D}}_\phi = \mathcal{D}_\phi \qquad y \qquad \mathbb{R}_> \subset im\ \nabla\phi \subset \mathbb{R}_+.$$

Sea $t_\phi = \sup_{t \in \mathcal{D}_\phi} t$, $\quad t^0_\phi = \sup_{\nabla\phi(t)=0} t$.
Definimos las siguientes familias de funciones:

$\Phi_s = \{\phi \in \Phi : \phi$ *es estrictamente convexa*$\}$,
$\Phi_0 = \{\phi \in \Phi : \phi$ *es estrictamente convexa en* $(t^0_\phi, t_\phi),\ t^0_\phi > \infty\}$,
$\Phi_{s'} = \{\phi \in \Phi_s : \inf \phi > -\infty\}$, $\Phi_{s''} = \{\phi \in \Phi_s : \inf \phi = -\infty\}$ *y* $\tilde{\Phi} = \Phi_0 \cup \Phi_s$.

Podemos ver que si $\phi \in \Phi$, entonces ϕ es no decreciente, ya que im $\nabla\phi \subset \mathbb{R}_+$. El dominio de ϕ viene dado por $\mathcal{D}_\phi = (-\infty, t_\phi)$.
Notemos que $t^0_\phi = -\infty$ si sólo si im $\nabla\phi = \mathbb{R}_>$, en efecto, si $t^0_\phi = -\infty$, entonces no existe $t \in \mathbb{R}$ tal que $\nabla\phi(t) = 0$, por lo cual $\mathbb{R}_> \subset$ im $\nabla\phi \subset \mathbb{R}_>$, y así $\mathbb{R}_> =$ im $\nabla\phi$. Recíprocamente si $\mathbb{R}_> =$ im $\nabla\phi$, entonces $t^0_\phi = \sup_{\nabla\phi(t)=0} t = -\infty$.
Otra observación interesante es la siguiente: $\phi \in \Phi_s$ si sólo si $\nabla\phi$ es creciente, (ver [16], Corolario 26.3.1), en efecto, si $\phi \in \Phi_s$, entonces ϕ es estrictamente convexa y como $\mathbb{R}_> \subset$ im $\nabla\phi \subset \mathbb{R}_+$, y así $\nabla\phi$ es creciente. De igual modo podemos decir que $\phi \in \Phi_0$ si sólo si $\nabla\phi$ es creciente en (t^0_ϕ, t_ϕ) y $t^0_\phi > -\infty$.
Otro resultado que podemos comentar es el siguiente: $\phi \in \Phi$ si sólo si ϕ es cerrada, propia, convexa y se verifica $\mathcal{D}_{\nabla\phi} = \overset{\circ}{\mathcal{D}_\phi} = \mathcal{D}_\phi$ y $\mathbb{R}_> \subset$ im $\nabla\phi \subset \mathbb{R}_+$. Lo cual es fácil de verificar usando la Proposición 1.2.7, obteniéndose que ϕ es esencialmente suave si sólo si $\mathcal{D}_{\nabla\phi} = \overset{\circ}{\mathcal{D}_\phi}$.
El dominio de la B-función asociada a la función de penalidad que usaremos para aplicar el método de multiplicadores basados en shifts, la cual es definida en (3.15), es $\mathcal{D}_{\theta^*} = [0, 1]$, por lo cual no se verifica la condición $\mathcal{D}_{\theta^*} \supset \mathbb{R}_>$, por consiguiente no pertenece a Ψ. De igual modo la función de penalidad θ no está en Φ, ya que im$\nabla\theta = [0, 1]$, de donde no se verifica $\mathbb{R}_> \subset$ im $\nabla\theta \subset \mathbb{R}_+$. En conclusión ni θ, ni θ', pertenecen a ninguna de las familias presentadas por Kiwiel en [9].

En las siguientes proposiciones se discute la relación de dualidad entre las familias de penalidad y las familias de funciones de Bregman.

Proposición 3.4.1 *(Ver [9], Lema 9.5)*
Si $\phi \in \Phi$, *entonces* ϕ^* *es una B-función y se verifica lo siguiente:*
$\mathbb{R}_> \subset \mathcal{D}_{\phi^*} \subset \mathbb{R}_+$, $(\phi^*)^+ = \phi^{**} = \phi$, $\lim_{t\downarrow-\infty} \nabla\phi(t) = 0$, $\lim_{t\uparrow t_\phi} \nabla\phi(t) =$

$\lim_{t\uparrow t_\phi}\phi(t)=\infty$ *y* $\phi'_\infty=\iota^*_{\mathbb{R}_+}$.
Si $\phi\in\Phi_s$, *entonces* ϕ^* *es esencialmente suave con:* $\nabla\phi^*=(\nabla\phi)^{-1}$ *y* $\mathcal{D}_{\partial\phi^*}=\mathcal{D}_{\nabla\phi^*}=\mathbb{R}_>$.
Si $\phi\in\Phi_0$, *entonces* $\nabla\phi^*=(\nabla\phi)^{-1}$, $\mathcal{D}_{\nabla\phi^*}=\mathbb{R}_>$ *y* $\partial\phi^*(0)=(-\infty,t^0_\phi]$.

Demostración . Si $\phi\in\Phi$, entonces es cerrada, propia, convexa, esencialmente suave y se verifica $\overset{\circ}{\mathcal{D}}_\phi=\mathcal{D}_\phi$ y $\mathbb{R}_>\subset\operatorname{im}\nabla\phi\subset\mathbb{R}_+$. Usando el Lema 3.1.3 se obtiene ϕ^* es una B-función con ri $\mathcal{D}_{\phi^*}\subset\operatorname{im}\nabla\phi\subset\mathcal{D}_{\phi^*}$. De donde se deduce que $\mathbb{R}_>\subset\mathcal{D}_{\phi^*}\subset\mathbb{R}_+$, por lo cual $\phi^*=\phi^*_+$ y así $(\phi^*)^+=(\phi^*_+)^*=\phi^{**}=\phi$.
Como $\mathbb{R}_>\subset\operatorname{im}\nabla\phi\subset\mathbb{R}_+$ y $\nabla\phi$ es no decreciente se tiene que $\lim_{t\downarrow-\infty}\nabla\phi(t)=0$ y $\lim_{t\uparrow t_\phi}\nabla\phi(t)=\infty$. Como ϕ es cerrada y propia usamos ([16], Teorema 13.3) para obtener $\phi'_\infty=\iota^*_{\mathcal{D}_{\phi^*}}$ con $\iota^*_{\mathcal{D}_{\phi^*}}=\iota^*_{\mathrm{cl}\ \mathcal{D}_{\phi^*}}$, además cl $\mathcal{D}_{\phi^*}=\mathbb{R}_+$, ya que $\mathbb{R}_>\subset\mathcal{D}_{\phi^*}\subset\mathbb{R}_+$. Por lo cual $\phi'_\infty=\iota^*_{\mathbb{R}_+}$.
Tomamos $t_\phi<\infty$, entonces $\lim_{t\uparrow t_\phi}\phi(t)=\infty$, ya que ϕ es cerrada y $t_\phi\notin\mathcal{D}_\phi$.
Si por el contrario $t_\phi=\infty$, se tiene $\lim_{t\uparrow t_\phi}\phi(t)=\infty$, ya que por ([16], Teorema 8.5) se tiene $\phi'_\infty(1)=\lim_{t\uparrow\infty}[\phi(t)-\phi(0)]/t=\infty$.
Si $\phi\in\Phi_s$, entonces es estrictamente convexa y por la Proposición 1.2.7 se deduce que ϕ^* es esencialmente suave y $\mathcal{D}_{\nabla\phi^*}=\mathcal{D}_{\partial\phi^*}=\overset{\circ}{\mathcal{D}}_{\phi^*}=\mathbb{R}_>$, por la Proposición 1.2.5 se tiene que $\nabla\phi^*=(\nabla\phi)^{-1}$.
Si $\phi\in\Phi_0$, entonces por Las Proposiciones 1.2.5 y 1.2.7 se tiene que $\partial\phi^*=(\nabla\phi)^{-1}=\{(\nabla\phi)^{-1}\}$. Por lo cual $\partial\phi^*(0)=\{t:\nabla\phi(t)=0\}=(-\infty,t^0_\phi]$, ya que $0\le\nabla\phi(t)\le\nabla\phi(t^0_\phi)$, $\forall t\le t^0_\phi$. Por otro lado por ser $\nabla\phi$ creciente en (t^0_ϕ,t_ϕ) obtenemos que $\partial\phi^*=\{(\nabla\phi)^{-1}\}$ es univaluada en $\mathbb{R}_>=(\nabla\phi(t^0_\phi),\infty)\subset\operatorname{im}\nabla\phi$ y así $\partial\phi^*=\{\nabla\phi^*\}$ en $\mathbb{R}_>$ (ver Proposición 1.2.6). ∎

Proposición 3.4.2 *(Ver [9], Lema 9.6)*
Sea ψ *una B-función en* $\mathbb{R}$ *tal que* $\mathcal{D}_\psi\supset\mathbb{R}_>$, *entonces* $\psi^+\in\Phi$. *Supongamos que* $\mathcal{D}_{\nabla\psi}\supset\mathbb{R}_>$. *Si* $\partial\psi(0)=\emptyset$ $(0\notin\mathcal{D}_\psi$ *ó* $\psi'(0;1)=-\infty)$, *entonces* ψ_+ *es esencialmente suave y* $\psi^+\in\Phi_s$.
Si $\partial\psi(0)\neq\emptyset$ $(\psi'(0;1)>-\infty)$, *entonces* $\psi^+\in\Phi_0$ *con* $t^0_{\psi^+}=\psi'(0;1)$

y existe una B-función $\breve{\psi}$ tal que $\psi_+ = \breve{\psi}_+$, $\psi^+ = \breve{\psi}^+$, $\mathcal{D}_{\nabla\breve{\psi}} \supset \mathbb{R}_+$ y $\nabla\breve{\psi}(0) = t^0_{\psi^+}$.

Demostración . Como ψ es una B-función y $\psi_+ = \psi + \iota_{\mathbb{R}_+}$, por el Lema 3.1.1 podemos ver que ψ_+ es una B-función. Además como $\psi^+ = \psi^*_+$ se deduce que $\mathcal{D}_{\psi^+} = \mathcal{D}_{\psi^*_+}$ y usando el Lema 3.1.3 se obtiene $\mathcal{D}_{\psi^+} = \overset{\circ}{\mathcal{D}}_{\psi^+}$.
Por otro lado ψ^+ es no decreciente y por el Lema 3.3.1 es esencialmente suave, por lo cual $\operatorname{im} \nabla\psi^+ \subset \mathbb{R}_+$. Como $\mathbb{R}_> \subset \mathcal{D}_\psi$ se obtiene que $\mathbb{R}_> \subset \mathcal{D}_{\psi_+}$, por lo cual $\mathbb{R}_> \subset \overset{\circ}{\mathcal{D}}_{\psi_+} \subset \mathcal{D}_{\psi_+} \subset \operatorname{im} \partial\psi^+ = \operatorname{im} \nabla\psi^+$, por lo tanto $\mathbb{R}_> \subset \operatorname{im} \nabla\psi^+ \subset \mathbb{R}_+$, y así $\psi^+ \in \Phi$.
Supongamos que $\mathcal{D}_{\nabla\psi} \supset \mathbb{R}_>$, como ψ es estrictamente convexa, por ser una B-función, se deduce $\nabla\psi_+ = \nabla\psi$ es creciente en $\mathbb{R}_>$, por lo cual $\nabla\psi^+ = (\nabla\psi_+)^{-1}$ es creciente en (t^0, ∞), donde $t^0 = \lim_{t\downarrow 0} \nabla\psi(t)$, por tanto ψ^+ es estrictamente convexa en (t^0, ∞) (ver [16], Corolario 26.3.1).
Si $\partial\psi(0) = \emptyset$, entonces $t^0 = -\infty$, y así ψ^+ es estrictamente convexa en $\mathbb{R}$. Por la Proposición 1.2.7 se obtiene que ψ_+ es esencialmente suave.
Si $\partial\psi(0) \neq \emptyset$, tenemos que $t^0 = \psi'(0;1) = t^0_{\psi^+}$. Sea $\breve{\psi}(t) = \psi(t)$, $\forall t \geq 0$ y sea $\breve{\psi}(t)$ para $t \leq 0$, una función convexa cuadrática sujeta a $\breve{\psi}(0) = \psi(0)$, y $\breve{\psi}'(0;-1) = -\psi'(0;1)$, entonces $\breve{\psi}_+ = \psi_+$ y $\nabla\breve{\psi}(0) = t^0_{\psi^+}$.

∎

Las Proposiciones 3.4.1 y 3.4.2 generan las siguientes relaciones de dualidad entre las familias de penalidad y las funciones de Bregman:
$\Phi^* = \Psi_+$.
En efecto, como $\Psi_+ = \{\psi_+ : \psi \in \Psi\} = \{\psi \in \Psi : \psi_+ = \psi\}$. Si $\psi \in \Psi_+$, entonces usando la Proposición 3.4.2 se obtiene $\psi^* = \psi^*_+ = \psi^+ \in \Phi$, por lo cual $\psi = \psi^{**} \in \Phi^*$.
Recíprocamente, si $\psi \in \Phi^*$, entonces $\psi^* \in \Phi$, por el Lema 3.1.3 ψ es una B-función y se cumple $\operatorname{ri} \mathcal{D}_\psi \subset \operatorname{im} \nabla\psi^* \subset \mathcal{D}_\psi$, pero por la Definición 3.4.2 $\mathbb{R}_> \subset \operatorname{im} \nabla\psi^* \subset \mathbb{R}_+$, por lo cual $\mathbb{R}_> \subset \mathcal{D}_\psi$. Por tanto $\psi \in \Psi$. Por otro lado como $\psi^* \in \Phi$, entonces por la Proposición 3.4.1 tenemos que $\mathbb{R}_> \subset \mathcal{D}_{\psi^{**}} = \mathcal{D}_\psi \subset \mathbb{R}_+$, por lo cual $\psi_+ = \psi$ y así $\psi \in \Psi_+$.
De manera similar podemos demostrar que $\tilde{\Phi}^* = \tilde{\Psi}_+$, $\Phi^*_s = \tilde{\Psi}_s$, $\Phi^*_{s'} = \tilde{\Psi}_{s'}$,

$\Phi^*_{s''} = \tilde{\Psi}_{s''}$, $\Phi^*_0 = \Psi_{0_+}$.
También tenemos que: $\Psi^+ = \Phi$, $\tilde{\Psi}^+ = \tilde{\Phi}$, $\Psi^+_s = \Phi_s$, $\Psi^+_{s'} = \Phi_{s'}$, $\Psi^+_{s''} = \Phi_{s''}$, $\Psi^+_0 = \Phi_0$.
A continuación se presenta una definición que permite dar una caracterización de las funciones en Φ.

Definición 3.4.3 FORCING. *(Ver [9], Definición 9.12)*
Sea $\phi \in \Phi$, decimos que ϕ es forcing en $[t'_\phi, t''_\phi]$ si

$$[\phi'(t'_k) - \phi'(t''_k)](t'_k - t''_k) \to 0,$$

implica:

$$\phi'(t''_k)(t'_k - t''_k) \to 0,$$

para cada par de suceciones $\{t'_k\}, \{t''_k\} \subset [t'_\phi, t''_\phi] \cap \mathcal{D}_\phi$, donde $\phi' = \nabla\phi$.

El siguiente lema establece que una función de penalidad acotada inferiormente y perteneciente a la familia Φ_s es forcing.

Lema 3.4.3 *Si $\phi \in \Phi_s$, $\inf \phi > -\infty$ y $t''_\phi \in \mathcal{D}_\phi$, entonces ϕ es forcing en $(-\infty, t''_\phi]$.*

Demostración . Reemplazando a ϕ por $\phi - \inf \phi$, obtenemos que $\inf \phi = 0$.
Por otro lado $\phi' = \nabla\phi$ es positiva y creciente, asi como lo es ϕ.
Si $[\phi'(t'_k) - \phi'(t_k)]\tau_k \to 0$, con $\tau_k > 0$, $t'_k = t_k + \tau_k \leq t''_\phi$.
Supongamos que $\phi'(t_k)\tau_k \not\to 0$, entonces existe $\epsilon > 0$ y $K = \{1, 2, \ldots\}$ tal que $\phi'(t_k)\tau_k \geq \epsilon$, $\forall k \in K$, por lo cual $\frac{\phi'(t'_k)}{\phi'(t_k)} \xrightarrow{K} 1$.
Como $\phi'(t_k) < \phi'(t''_\phi)$ tenemos que $\tau_k \geq \frac{\epsilon}{\phi'(t''_\phi)}$, $\forall k \in K$. Por otro lado $\phi(t'_k) \geq \phi(t_k) + \phi'(t_k)\tau_k \geq \epsilon$, por lo cual $t'_k \geq \phi^{-1}(\epsilon)$, $\forall k \in K$.
Elegimos t_∞ y $K' \subset K$, tal que $t'_k \xrightarrow{K'} t_\infty$. Entonces $t_k + \frac{\epsilon}{2\phi'(t''_\phi)} \leq t_\infty$ y $\phi'(t_k) \leq \phi'(t_k - \frac{\epsilon}{2\phi'(t''_\phi)}) < \phi'(t_\infty) = \lim_{k \in K'} \phi'(t'_k)$, para un $k \in K'$ suficientemente grande, lo cual contradice $\frac{\phi'(t'_k)}{\phi'(t_k)} \xrightarrow{K} 1$. Por tanto $\phi'(t_k)\tau_k \to 0$, de donde ϕ es forcing. ∎

Capítulo 4

IMPLEMENTACION DEL MÉTODO DE MULTIPLICADORES BASADOS EN SHIFTS.

En este capítulo se implementará el método de multiplicadores basados en Shifts con una penalidad no coerciva, este método es considerado por Castillo en [3] y lo aplica para dos familias de penalidades, llamadas AL1 y AL2. La familia AL1 cumple con algunas propiedades que ya estudiamos en el capítulo 2, entre las cuales destaca que no son coercivas. Por el contrario la familia de penalidades AL2 es una familia de funciones coercivas.

En este trabajo usaremos particularmente la función definida en (3.10), la cual es considerada por Zang en [19] en problemas de minimización como una aproximación de la función

$$t \in \mathbb{R} \mapsto q(t) = \max\{0, t\}$$

Acerca de esta función estudiaremos algunas de sus propiedades, así como la de su conjugada.

Se presentan en este capítulo varios resultados relacionados con este método en consideración, los cuales son presentados por Castillo en [3]. Se hace una demostración detallada de cada uno de ellos.

El aporte principal de este trabajo se presenta en este capítulo, el cual consiste en establecer una demostración de la convergencia de la sucesión $\{x^k\}$ generada por el método de multiplicadores basados en shifts con la penalidad no coerciva θ. Un resultado similar muestra Castillo en [3] para las familias AL1 y AL2, no obstante realiza la demostración para AL2 y deja plantedo el caso para AL1. La demostración que se realiza en este trabajo corresponde al caso particular de la penalidad θ, la cual es similar a las penalidades de tipo AL1, en el sentido que cumple todas las propiedades de AL1, excepto la convexidad estricta. Entonces la demostración realizada en este trabajo es similar a la que deja propuesta Castillo en [9]. Además la convergencia se establece sin necesidad de imponer dos condiciones que establece Castillo en su demostración.

Otro aporte que se hace en este capítulo es una implementación numérica, con el fin de mostrar el desempeño computacional de este método de multiplicadores basados en shifts con la penalidad θ, para la cual se efectuó un programa con el sofware Matlab, el cual aplica el algoritmo del método de multiplicadores basados en shifts con la penalidad no coerciva θ. Se aplica este programa para resolver algunos problemas propuestos en [7] y [2], obteniendo como salida los valores aproximados de los minimizadores de la función objetivo f, así como los maximizadores de la función objetivo dual.

4-1 ESTUDIO DE LA FUNCIÓN DE PENALIDAD θ-

Consideremos la función definida en (3.10):

$$\theta(x) = \begin{cases} 0 & \text{si} \quad x \leq -1 \\ b(x) & \text{si} \quad -1 \leq x \leq 1 \\ x & \text{si} \quad x \geq 1, \end{cases} \tag{4.1}$$

donde

$$b(x) = \frac{1}{4}x^2 + \frac{1}{2}x + \frac{1}{4} = \frac{1}{4}(x+1)^2.$$

Figura 4.1: Gráfica de la función de penalidad θ.

En primer lugar debemos notar que $\mathcal{D}_\theta = \mathbb{R}$, por lo cual tenemos que θ es propia, debido a que $\theta(x) < \infty, \quad \forall x \in \mathbb{R}$.

La función θ es continua en $\mathbb{R}$, por lo cual $\liminf_{y \to x} \theta(y) = \theta(x)$, de donde θ es cerrada.

También tenemos que θ es convexa y no decreciente.

Con respecto a θ podemos decir que cumple las condiciones de la familia AL1, excepto la convexidad estricta.
En efecto:

i) θ no es estrictamente convexa, ya que si tomamos $x_1, x_2 \geq 1$ y $\alpha \in (0,1)$, tenemos $\alpha x_1 + (1-\alpha)x_2 \geq 1$, por lo cual:

$$\theta(\alpha x_1 + (1-\alpha)x_2) = \alpha x_1 + (1-\alpha)x_2 = \alpha\theta(x_1) + (1-\alpha)\theta(x_2).$$

ii) θ es diferenciable en $\mathbb{R}$. Como la función nula, la función identidad y la función b son diferenciables, es suficiente probar que θ es diferenciable en -1 y en 1.
Como

$$\begin{aligned}
\lim_{\lambda \downarrow 0} \frac{\theta(-1+\lambda) - \theta(-1)}{\lambda} &= \lim_{\lambda \downarrow 0} \frac{\frac{1}{4}(-1+\lambda)^2 + \frac{1}{2}(-1+\lambda) + \frac{1}{4} - 0}{\lambda} \\
&= \lim_{\lambda \downarrow 0} \frac{\frac{1}{4}(1^2 - 2\lambda + \lambda^2) - \frac{1}{2} + \frac{1}{2}\lambda + \frac{1}{4}}{\lambda} \\
&= \lim_{\lambda \downarrow 0} \frac{\lambda^2}{4\lambda} = 0.
\end{aligned}$$

Luego θ es diferenciable en -1. De manera similar se prueba que θ es diferenciable en 1.
Podemos considerar la función derivada $\theta^{'} : \mathbb{R} \to \mathbb{R}$, definida de la siguiente manera:

$$\theta^{'}(x) = \begin{cases} 0 & \text{si} \quad x \leq -1 \\ \frac{1}{2}(x+1) & \text{si} \quad -1 \leq x \leq 1 \\ 1 & \text{si} \quad x \geq 1. \end{cases} \tag{4.2}$$

Debemos notar que:

$$0 \leq \theta^{'}(x) \leq 1, \quad \forall x \in \mathbb{R}. \tag{4.3}$$

Además en el intervalo $[0,1]$ podemos deteminar la inversa de $\theta^{'}$, la cual está definida por:

$$(\theta^{'})^{-1}(x) = 2x - 1, \tag{4.4}$$

Figura 4.2: Gráfica de la derivada θ'.

por lo cual:

$$-1 \leq (\theta')^{-1}(x) \leq 1, \quad \forall x \in [0,1]. \tag{4.5}$$

iii) $\theta'_\infty(1) = 1$, aplicando la Definición 1.1.17 obtenemos:

$$\theta_\infty{}'(1) = \lim_{\lambda \to +\infty} \frac{\theta(x+\lambda) - \theta(x)}{\lambda}.$$

Como $\theta(y) = y, \ \forall y \geq 1$. Podemos tomar entonces un $\lambda^* > 0$ tal que $\theta(x+\lambda) = x + \lambda, \quad \forall \lambda \geq \lambda^*$.

Luego para cada $\lambda \geq \lambda^*$ se verifica

$$\frac{\theta(x+\lambda) - \theta(x)}{\lambda} = \frac{x + \lambda - \theta(x)}{\lambda} = \frac{x}{\lambda} + 1 - \frac{\theta(x)}{\lambda}.$$

Tomando límite obtenemos:

$$\theta_\infty{}'(1) = \lim_{\lambda \to +\infty} \frac{\theta(x+\lambda) - \theta(x)}{\lambda} = 1.$$

Por esta propiedad de θ podemos decir que es no coerciva. (ver Definición 1.1.20)

Esta propiedad es de sumo interés, ya que las funciones de penalidad consideradas en [9] son coercivas y por lo tanto los resultados y el análisis obtenidos en su investigación se aplican a estas funciones. Surge de esta manera la necesidad de estudiar el funcionamiento del método de multiplicadores basados en shifts con esta penalidad no coerciva.

iv) $\lim_{y\to-\infty}\theta'(y)=0$, ya que $\theta'(y)=0, \quad \forall y\leq -1$.

v) θ es acotada inferiormente, en efecto $\theta(x)\geq 0, \quad \forall x\in\mathbb{R}$.

vi) $\sup_{x>0}\{x-\theta(x)\}$ es finito.
Tenemos que $x-\theta(x)=0, \quad \forall x\geq 1$.
Por otro lado si definimos la función $g(x)=x-\frac{1}{4}x^2-\frac{1}{2}x-\frac{1}{4}$ en el intervalo $[0,1]$.
Derivando obtenemos: $g'(x)=-\frac{1}{2}x+\frac{1}{2}$.
Luego $g'(x)=0$ cuando $x=1$, además $g''(1)=-\frac{1}{2}<0$.
Por lo cual $\sup_{x>0}\{x-\theta(x)\}=1-\theta(1)=0<\infty$.

Como θ cumple las condiciones de AL1, exceptuando la convexidad estricta, pueden usarse algunos resultados que no requieran esta condición expresados en [3].

Como θ es diferenciable en $\mathbb{R}$, se tiene que es esencialmente suave, ya que $\overset{\circ}{\mathcal{D}}_\theta=\mathbb{R}$. Usando la Proposición 1.2.7(i), se tiene que $\partial\theta(x)=\nabla\theta(x), \quad \forall x\in\mathbb{R}$.
Por (3.15) la conjugada de θ es la función $\theta^*[0,1]\to\mathbb{R}$ definida por:

$$\theta^*(y)=y(y-1). \tag{4.6}$$

Por La Proposición 1.2.7(ii) se deduce que θ^* es esencialmente estricta convexa en [0,1], ya que θ es esencialmente suave y por ser cerrada, se tiene que $\theta^{**}=\theta$.
Usando el Lema 3.1.3(ii) tenemos que θ^* es una B-función y se verifica:

$$\text{ri }\mathcal{D}_{\theta^*}\subset \text{im }\nabla\theta\subset\mathcal{D}_{\theta^*}.$$

Notemos que ri $\mathcal{D}_{\theta^*}=(0,1)$, im $\nabla\theta=[0,1]$ y $\mathcal{D}_{\theta^*}=[0,1]$.
Una diferencia notable de la función θ^*, con las funciones de Bregman estudiadas en [9], es que $\mathbb{R}_>\not\subset\mathcal{D}_{\theta^*}$, y así θ no es considerada como función de penalidad en ninguno de los análisis allí presentados.

Figura 4.3: Gráfica de la función de Bregman θ^*.

Asociada a la función θ^*, podemos considerar la Distancia de Bregman $D_{\theta^*}(x,y)$ de la siguiente manera:

$$\begin{aligned} D_{\theta^*}(x,y) &= \theta^*(x) - \theta^*(y) - \nabla\theta^*(y)(x-y) \\ &= x(x-1) - y(y-1) - (2y-1)(x-y) \\ &= x(x-1) - y(y-1) - (y-1+y)(x-y) \\ &= x(x-1) + (y-1)(y-x-y) - y(x-y) \\ &= x(x-1-y+1) - y(x-y) \\ &= (x-y)^2. \end{aligned} \tag{4.7}$$

Podemos también calcular el gradiente:

$$\nabla_x D_{\theta^*}(x,y) = 2(x-y). \tag{4.8}$$

Además:

$$\nabla_y D_{\theta^*}(x,y) = -2(x-y). \tag{4.9}$$

Consideremos el método de multiplicadores basados en shifts según el Algoritmo 2.1.1, tomando en cuenta el multiplicador definido en (2.13) de la manera siguiente:

$$x \in \mathbb{R}^n, \mu \in (0,\beta) \mapsto L_{\beta,r}(x,\mu) = f(x)+\sum_{i=1}^{m} \beta_i r \left[\theta\left(\frac{f_i(x)}{r} + \tilde{y}_i(\mu_i)\right) - \theta(\tilde{y}_i(\mu_i))\right] \tag{4.10}$$

donde la función de penalidad θ es la definida en (4.1).

Lema 4.1.1 *(Ver Proposición 2.1, [1])*
Sea la función θ *definida en (4.1), sea* f *una función propia, cerrada y convexa tal que* $\mathcal{D}_\theta \cap f(\mathbb{R}^n) \neq \emptyset$ *y consideremos la función compuesta:*

$$g(x) = \begin{cases} \theta(f(x)) & si \quad x \in \mathcal{D}_f \\ +\infty & de\ otra\ forma, \end{cases}$$

entonces g *es una función propia, cerrada y convexa y se verifica:*

$$g'_\infty(d) = \begin{cases} \theta'_\infty(f'_\infty(d)) & si \quad d \in \mathcal{D}_{f'_\infty} \\ +\infty & de\ otra\ forma. \end{cases}$$

Demostración . Como $\mathcal{D}_\theta \cap f(\mathbb{R}^n) \neq \emptyset$, se tiene que existe un $x \in \mathbb{R}^n$ tal que $g(x) < \infty$, luego g es propia.
Como θ es continua y f es cerrada se tiene que g es cerrada.
La función g es convexa debido a la convexidad de f y por ser θ no decreciente.
Sea $x \in \mathcal{D}_f$ tal que $f(x) \in \mathcal{D}_\theta$. Para cada $a < f'_\infty(d)$ existe un $\bar{t}$ tal que:

$$f(x+td) \geq f(x) + ta, \quad \forall t \geq \bar{t}.$$

Además como θ es no decreciente se tiene:

$$\frac{g(x+td) - g(x)}{t} \geq \frac{\theta(f(x)+ta) - \theta(f(x))}{t}.$$

Tomando límite cuando $t \to +\infty$ obtenemos que $g'_\infty(d) \geq \theta'_\infty(a)$.
Si $f'_\infty(d) = +\infty$, considerando que $\theta'_\infty(1) = 1 > 0$, y haciendo que $a \to \infty$ se obtiene $g'_\infty(d) = +\infty$.
Si $f'_\infty(d) < +\infty$ tomamos $a \to f'_\infty(d)$, y como θ_∞ es semicontinua inferior se deduce $g'_\infty(d) \geq \theta'_\infty(f'_\infty(d))$.

Por otro lado tenemos que $f(x+td) \leq f(x) + tf'_\infty(d)$. Usando el hecho que θ es no decreciente, deducimos:

$$\begin{aligned} g'_\infty(d) &= \lim_{t\to+\infty} \frac{g(x+td)-g(x)}{t} \\ &\leq \lim_{t\to+\infty} \frac{\theta(f(x)+tf'_\infty(d))-\theta(f(x))}{t} \\ &= \theta'_\infty(f_\infty(d)). \end{aligned}$$

Por lo tanto concluimos que $g'_\infty(d) = \theta'_\infty(f'_\infty(d))$.

■

A continuación presentamos algunos resultados presentados por Castillo en [3], los cuales muestran algunos aspectos importantes de la penalidad usada en los métodos de multiplicadores basados en shifts, en este trabajo presentamos estos resultados usando de manera particular la función de penalidad θ.

Lema 4.1.2 *(Ver [3],Lema 5.2)*
Sea la función θ definida en (4.1), Sea $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ una función cerrada y convexa. Sean $r \in (0,1]$, $\beta \geq 1$, $\mu > 0$ fijos y $\tilde{y} \in \mathbb{R}$ tal que $\theta'(\tilde{y}) = \frac{\mu}{\beta} < \frac{1}{2}$.
Consideremos:

$$p(x) = \begin{cases} \beta r \left[\theta\left(\frac{f(x)}{r} + \tilde{y}\right) + \theta(\tilde{y})\right] & \textit{si} \quad x \in \mathcal{D}_f \\ \infty & \textit{caso contrario,} \end{cases} \tag{4.11}$$

entonces $\forall d \in \mathbb{R}^n,\ d \neq 0$, se cumple:

$$p'_\infty(x) = \begin{cases} \beta\theta'_\infty(f'_\infty(d)) & \textit{si} \quad d \in \mathcal{D}_{f'_\infty} \\ \infty & \textit{caso contrario.} \end{cases} \tag{4.12}$$

Demostración . Para β, μ y r fijos escribimos p de la siguiente forma:

$$p(x) = \begin{cases} \beta r\left[\theta(h(x)) + C\right] & \text{si} \quad x \in \mathcal{D}_f \\ \infty & \text{caso contrario,} \end{cases} \tag{4.13}$$

donde $h(x) = \frac{f(x)}{r} + \tilde{y}$ y $C = \beta r = \theta(\tilde{y})$.
De esta forma tenemos que h es propia, cerrada y convexa, además $\mathcal{D}_h = \mathcal{D}_f$.
Tenemos que:

$$
\begin{aligned}
h'_\infty(d) &= \lim_{\lambda\to+\infty} \frac{h(x+\lambda d) - h(x)}{\lambda} \\
&= \lim_{\lambda\to+\infty} \frac{\frac{f(x+\lambda d)}{r} + \tilde{y} - \frac{f(x)}{r} - \tilde{y}}{\lambda} \\
&= \frac{1}{r} \lim_{\lambda\to+\infty} \frac{f(x+\lambda d) - f(x)}{\lambda} \\
&= \frac{f'_\infty(d)}{r}.
\end{aligned}
$$

Usando el lema anterior nos queda:

$$
\begin{aligned}
p'_\infty(d) &= \begin{cases} \beta r \theta'_\infty(h'_\infty(d)) & \text{si} \quad d \in \mathcal{D}_{h'_\infty} \\ +\infty & \text{de otra forma.} \end{cases} \\
&= \begin{cases} \beta \theta'_\infty(f'_\infty(d)) & \text{si} \quad d \in \mathcal{D}_{f'_\infty} \\ +\infty & \text{de otra forma.} \end{cases}
\end{aligned}
$$

∎

Para el problema (P) primal podemos considerar las siguientes hipótesis:

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

(A2) Existe $\bar{x} \in \mathcal{D}_f$ tal que $f_i(\bar{x}) < 0$, $i = 1, \ldots, m$. Esta condición es llamada Condición de Slater.

Proposición 4.1.3 *(Ver [3], Lema 5.4)*
Sea la función θ definida en (4.1), sea $r \in (0,1]$ fijo. Supongamos que se satisface (A1), entonces existe $\bar{\beta} \geq e = (1,1,\ldots,1)$ tal que para todo $\beta > \bar{\beta} \geq e$ y para μ con $\mu_i \in (0, \beta_i)$, el conjunto

$$S = \arg\min_x L_{\beta,r}(x,\mu),$$

es no vacío y compacto. Si $(f_0)'_\infty$ es no negativa, entonces la conclusión es válida para todo $\beta \geq e$.

Demostración .

Tomemos $r \in (0,1]$ fijo. Sea μ tal que $\mu_i \in (0, \beta_i)$ para $i = 1, \ldots, m$, para facilitar la notación escribiremos que $\mu \in (0, \beta)$.

Consideremos $\tilde{y}_i$ tal que $\theta'(\tilde{y}_i) = \mu_i / \beta_i \leq 1$.

Para facilitar la notación escribimos:

$$L(x) = L_{\beta,r}(x,\mu) = f_0(x) + \sum_{i=1}^{m} \beta_i r \left[\theta\left(\frac{f_i(x)}{r} + \tilde{y}_i \right) - \theta(\tilde{y}_i) \right].$$

Consideremos $\forall d \in \mathbb{R}^n$, $d \neq 0$ a $I^+_\infty(d)$ el conjunto de indices $i = 1, \ldots, m$ tales que $(f_i)'_\infty > 0$. Del mismo modo definimos $I^-_\infty(d) = \{i : (f_i)'_\infty \leq 0\}$. Además denotemos $\mathcal{K} = \{d \in \mathbb{R}^n : d \in \bigcap_{i=1}^{m} \mathcal{D}_{(f_i)'_\infty}\}$.

Usando las propiedades de la función de recesión (ver [16], Teorema 9.3), así como el Lema 4.1.1, tenemos:

$$L'_\infty(d) = \begin{cases} (f_0)'_\infty(d) + \sum_{i=1}^{m} \beta_i \theta'_\infty((f_i)'_\infty(d)) & \text{si} \quad d \in \mathcal{K} \\ \\ +\infty & \text{de otra forma.} \end{cases}$$

Tomamos $d \in \mathcal{K}$ tal que $\|d\| = 1$, entonces para cada $\beta \geq e$ y $\mu \in (0, \beta)$, se deduce:

$$\sum_{i=1}^{m} \beta_i \theta'_\infty((f_i)'_\infty(d)) = \sum_{i \in I^+_\infty(d)} \beta_i (f_i)'_\infty(d) \theta'_\infty(1) + \sum_{i \in I^-_\infty(d)} \beta_i (f_i)'_\infty(d) \theta'_\infty(-1).$$

Como $\theta'_\infty(1) = 1$ y $\theta'_\infty(-1) = 0$, tenemos que para $d \in \mathcal{K}$ tal que $\|d\| = 1$ se cumple:

$$L'_\infty(d) = \begin{cases} (f_0)'_\infty(d) & \text{si} \quad (f_i)'_\infty(d) \leq 0, \ \forall i. \\ (f_0)'_\infty(d) + \sum_{i \in I^+_\infty(d)} \beta_i (f_i)'_\infty(d) & \text{si} \quad I^+_\infty(d) \neq \emptyset \\ \\ +\infty & \text{si} \quad d \notin \mathcal{K}. \end{cases}$$

Estudiamos cada uno de los casos: si $L'_\infty(d) = (f_0)'_\infty(d)$, se tiene que $L'_\infty(d) > 0$, ya que $(f_0)'_\infty(d) > 0$, por que de lo contrario se contradice la hipótesis (A1).

Si $L'_{\infty}(d) = (f_0)'_{\infty}(d) + \sum_{i \in I^+_{\infty}(d)} \beta_i (f_i)'_{\infty}(d)$, entonces: $L'_{\infty}(d) > 0$, ya que:

$$(f_0)'_{\infty}(d) > 0 \quad \text{y} \quad \sum_{i \in I^+_{\infty}(d)} \beta_i (f_i)'_{\infty}(d) > 0.$$

Por lo cual:

$$L'_{\infty}(\cdot) = (L_{\beta,r})'_{\infty}(\cdot) > 0, \quad \forall \beta \geq e.$$

Lo cual quiere decir que S es no vacío y compacto.
Si $(f_0)'_{\infty}(d) \leq 0$. Usando la continuidad de la función de recesión, deducimos que existe un $\delta > 0$ tal que:

$$\forall d \in \mathcal{K}, \quad \|d\| = 1, \quad \max_{i=1,\ldots,m} (f_i)'_{\infty}(d) \geq \delta.$$

De esta forma tenemos para $\beta \geq e$, $\mu \in (0, \beta)$ y $\forall d \in \mathcal{K}$, $\|d\| = 1$,

$$\sum_{i=1}^{m} \beta_i \theta'_{\infty}((f_i)'_{\infty}(d)) \geq \beta_j \delta, \quad \text{para algún} \quad j \in \{1, \ldots, m\}.$$

Por lo cual, $\forall d \in \mathcal{K}$, $\|d\| = 1$,

$$L'_{\infty}(d) \geq (f_0)'_{\infty}(d) + \beta_j \delta.$$

Si tomamos para cada $j = 1, \ldots, m$

$$\bar{\beta}_j \geq \frac{-\inf\{(f_0)'_{\infty}(d) : d \in \mathcal{K}, \quad \|d\| = 1\}}{\delta},$$

obtenemos que $(f_0)'_{\infty}(d) + b\bar{e}ta_j \delta \geq 0$. Por tanto: $\forall \beta > \bar{\beta}$, $\forall \mu \in (0, \beta)$, $\forall d \in \mathcal{K}$, $\|d\| = 1$, se verifica:

$$L'_{\infty}(d) = (L_{\beta,r})'_{\infty}(d) > 0.$$

∎

La proposición anterior garantiza que las minimizaciones consideradas en el algoritmo 2.1.1 están bien definidas cuando iniciamos con un $\beta^0 \geq \bar{\beta}$.

Proposición 4.1.4 *(Ver [3], Proposición 5.6)*
Sea θ *la función de penalidad definida por (4.1). Dados* $\beta \in \mathbb{R}^m, \quad r \in (0,1]$ *y* $\mu \in (0,\beta)$. *Consideremos la función*

$$y \in \mathbb{R}^m,\ \mu \in (0,\beta) \mapsto P(y) = \sum_{i=1}^{m} p(y_i) = \sum_{i=1}^{m} \beta_i r \left[\theta\left(\frac{y_i}{r} + \tilde{y}_i\right) - \theta(\tilde{y}_i,\right]$$

donde $\theta'(\tilde{y}_i) = \mu_i/\beta_i < 1$ *para* $i = 1,\ldots,m$. *Entonces se cumple:*

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

Demostración . Si tomamos $f(y_i) \equiv \theta(y_i + \tilde{y}_i)$, entonces usando la Proposición 1.2.1(5) obtenemos: $f^*(s_i) = \theta^*(s_i) - y_i s_i$.
Si $h(y_i) \equiv \beta_i r f\left(\frac{y_i}{r}\right)$, entonces, usando la Proposición 1.2.1(2) se deduce:

$$\begin{aligned} h^*(s_i) &= \beta_i r f^*\left(\frac{s_i}{\beta_i}\right) \\ &= \beta_i r\left(\theta^*\left(\frac{s_i}{\beta_i}\right) - \tilde{y}_i\frac{s_i}{\beta_i}\right) \\ &= \beta_i r\theta^*\left(\frac{s_i}{\beta_i}\right) - r\tilde{y}_i s_i. \end{aligned}$$

Además $p(y_i) = h(y_i) - C_i$, donde $C_i = \beta_i r\theta(\tilde{y}_i)$, para $i = 1,\ldots,m$. Por lo cual

$$\begin{aligned} p^*(s_i) &= h^*(s_i) + C_i \\ &= r\left[\beta_i\theta^*\left(\frac{s_i}{\beta_i}\right) - \tilde{y}_i s_i + \beta_i\theta(\tilde{y}_i)\right]. \end{aligned}$$

Ahora usamos la Proposición 1.2.5, con $x = y_i$ y $s_i = \frac{\mu_i}{\beta_i}$, de donde se deduce:

$$\theta(\tilde{y}_i) = \frac{\mu_i}{\beta_i}\tilde{y}_i - \theta^*\left(\frac{\mu_i}{\beta_i}.\right)$$

Por tanto tenemos que para cada $i = 1,\ldots,m$, se tiene:

$$\begin{aligned} p^*(s_i) &= r\left[\beta_i\theta^*\left(\frac{s_i}{\beta_i}\right) - \tilde{y}_i s_i + \left(\mu_i\tilde{y}_i - \beta_i\theta^*\left(\frac{\mu_i}{\beta_i}\right)\right)\right] \\ &= r\left[\beta_i\theta^*\left(\frac{s_i}{\beta_i}\right) - \beta_i\theta^*\left(\frac{\mu_i}{\beta_i}\right) - \tilde{y}_i(s_i - \mu_i)\right] \\ &= r\left[\beta_i\theta^*\left(\frac{s_i}{\beta_i}\right) - \beta_i\theta^*\left(\frac{\mu_i}{\beta_i}\right) - (\theta^*)'\left(\frac{\mu_i}{\beta_i}\right)(s_i - \mu_i)\right]. \end{aligned}$$

Concluimos que:

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

∎

Consideremos una nueva versión del problema dual (D) escrito de la siguiente manera:

$$(\bar{D}) \qquad \begin{array}{ll} \text{maximizar} & \varphi(\mu) \\ \text{sujeto a} & \mu \in [0,\beta] \end{array}$$

con $\beta \geq e$ fijo.

El método de punto proximal se puede usar para resolver el problema $(\bar{D})$, usando la distancia de Bregman asociada a θ^*, genera una sucesión tal que:

$$\mu^0 \in S_1 = (0,\beta),$$

$$\mu^{k+1} \in \arg\min_{\mu\in\bar{S}_1}\{-\varphi(\mu) + r^k D_h(\mu,\mu^k)\}, \tag{4.14}$$

donde

$$s \in \bar{S}_1 \mapsto h(s) = \sum_{i=1}^{m}\beta_i\theta^*\left(\frac{s_i}{\beta_i}\right). \tag{4.15}$$

La expresión (4.14) es equivalente a:

$$\mu^{k+1} \in \arg\max_{\mu\in\bar{S}_1}\{\varphi(\mu) - r^k D_h(\mu,\mu^k)\}, \tag{4.16}$$

donde $r^k \in (0,1]$ para todo k. Notémos que las expresiones (4.14) y (4.16) están bien definidas, en virtud de la convexidad estricta de $D_{\theta^*}(\cdot,\mu)$ y por ser coerciva.

La función h genera la siguiente distancia de Bregman:

$$D_h(s,\mu) = \sum_{i=1}^{m}\left[\beta_i\theta^*\left(\frac{s_i}{\beta_i}\right) - \beta_i\theta^*\left(\frac{\mu_i}{\beta_i}\right) - (\theta^*)'\left(\frac{\mu_i}{\beta_i}\right)(s_i-\mu_i)\right]. \tag{4.17}$$

El gradiente de la función $D_h(s,\mu)$ viene dado por:

$$\nabla_s D_h(s,\mu) = \sum_{i=1}^{m}\left[(\theta^*)'\left(\frac{s_i}{\beta_i}\right) - (\theta^*)'\left(\frac{\mu_i}{\beta_i}\right)\right]. \tag{4.18}$$

Teorema 4.1.5 *(Ver [3], Teorema 5.15)*
Consideremos una sucesión $\{r^k\}$ en $(0,1]$, sea $\{\mu^k\}$ la sucesión generada por (4.14) aplicada al Problema $(\bar{D})$, y sean $\{\hat{x}^k\}$ y $\{\hat{\mu}^k\}$ las sucesiones por (2.14) y (2.15)generadas por el Algoritmo 2.1.1 correspondiente al método de multiplicadores basados en shifts con la penalidad θ. Si $\hat{\mu}^0 = \mu^0$ y $\beta = \beta^k$ para todo $k \in \mathbb{N}$, donde $\beta \in \mathbb{R}^m_>$, entonces $\forall k \in \mathbb{N}$, $\hat{\mu}^k = \mu^k$.

Demostración . Consideremos $\beta^k = \beta$ para todo $k \in \mathbb{N}$, donde $\beta \in \mathbb{R}^m_>$.
Usaremos inducción:
Por hipótesis $\hat{\mu}^0 = \mu^0$, supongamos que $\hat{\mu}^k = \mu^k$. para $k \in \mathbb{N}$.
Sea $\tilde{y}$ tal que $\theta(\tilde{y}_i) = \frac{\hat{\mu}_i^k}{\beta}$, Si $\hat{x}^{k+1} \in \arg\min_x L_{\beta,r^k}(\cdot, \hat{\mu}^k)$, entonces por (2.16) tenemos que $\hat{x}^{k+1}$ minimiza al lagrangeano clásico $L(\cdot, \hat{\mu}^k)$.
Por otra parte, para todo $\mu \in \mathbb{R}^m$

$$\begin{aligned}
\varphi(\mu) &= \inf\{f(x) + \sum_{i=1}^m \mu_i f_i(x) : x \in \mathbb{R}^n\} \\
&\leq f(\hat{x}^{k+1}) + \sum_{i=1}^m \mu_i f_i(\hat{x}^{k+1}) \\
&= f(\hat{x}^{k+1}) + \sum_{i=1}^m \hat{\mu}_i^{k+1} f_i(\hat{x}^{k+1}) + \sum_{i=1}^m (\mu_i - \hat{\mu}_i^{k+1}) f_i(\hat{x}^{k+1}) \\
&= \varphi(\hat{\mu}_i^{k+1}) + \sum_{i=1}^m (\mu_i - \hat{\mu}_i^{k+1}) f_i(\hat{x}^{k+1}).
\end{aligned}$$

Por lo cual $(f_1(\hat{x}^{k+1}), \ldots, f_n(\hat{x}^{k+1})) \in \partial\varphi(\hat{\mu}_i^{k+1})$, ya que φ es concáva.
Por (2.15) tenemos $\forall i = 1, \ldots, m$.

$$\begin{aligned}
\hat{\mu}_i^{k+1} = \beta_i \theta' \left(\frac{f_i(\hat{x}^{k+1})}{r^k} + \tilde{y}_i \right) &\iff \frac{\hat{\mu}_i^{k+1}}{\beta_i} = \theta' \left(\frac{f_i(\hat{x}^{k+1})}{r^k} + \tilde{y}_i \right) \\
&\iff (\theta')^{-1} \left(\frac{\hat{\mu}_i^{k+1}}{\beta_i} \right) - \tilde{y}_i = \frac{f_i(\hat{x}^{k+1})}{r^k} \\
&\iff f_i(\hat{x}^{k+1}) = r^k \left[(\theta')^{-1} \left(\frac{\hat{\mu}_i^{k+1}}{\beta_i} \right) - \tilde{y}_i \right] \\
&\iff f_i(\hat{x}^{k+1}) = r^k \left[(\theta')^{-1} \left(\frac{\hat{\mu}_i^{k+1}}{\beta_i} \right) - (\theta')^{-1} \left(\frac{\hat{\mu}_i^{k}}{\beta_i} \right) \right].
\end{aligned}$$

Usando la hipótesis de inducción, tenemos $\forall i = 1, \ldots, m$:

$$f_i(\hat{x}^{k+1}) = r^k \left[(\theta')^{-1} \left(\frac{\hat{\mu}_i^{k+1}}{\beta_i} \right) - (\theta')^{-1} \left(\frac{\mu_i^k}{\beta_i} \right) \right] \in \partial\varphi(\hat{\mu}_i^{k+1}). \tag{4.19}$$

Usando (4.18) y $(\theta')^{-1} = (\theta^*)'$, concluimos que $\hat{\mu}_i^{k+1}$ es una solución única de (4.14) obtenida a partir de μ^k, por lo cual $\hat{\mu}^{k+1} = \mu^{k+1}$. ∎

Proposición 4.1.6 *(Ver [3], Proposición 5.17)*
Sea $\{\mu^k\}$ la sucesión (4.16) generada por el método de punto proximal, entonces:

i) La sucesión $\{\mu^k\}$ es acotada.

ii) La sucesión $\{\beta^k\}$ generada por el algoritmo 2.1.1, correspondiente al método de multiplicadores basados en shifts con penalidad θ, es acotada. Es decir existe $k_1 \in \mathbb{N}$ tal que $\forall k \geq k_1$, $\beta^k = \beta^{k_1}$.

Demostración .

i) De (A2) (Condición de Slater) se deduce que la función $\varphi(\cdot)$ es acotada superiormente por un valor óptimo $\hat{f}$. Como h es una función de Bregman, el conjunto $D_h(\cdot, \mu^k)$ tiene conjunto de nivel acotado.
Definimos los conjuntos:

$$L = \{\mu \in S_1 : \varphi(\mu) \geq \varphi(\mu^0)\} \qquad Q = \max\{\|\mu\|_\infty : \mu \in L\},$$

de esta forma $\mu^k \in L$ para todo $k \in \mathbb{N}$. Por lo tanto $\|\mu^k\|_\infty \leq Q$, $\forall k \in \mathbb{N}$.

ii) La actualización de β^k en el algoritmo 2.1.1 establece que para $i = 1, \ldots, m$; si $\mu_i^k > \frac{\beta_i^k}{2}$, entonces $\beta_i^{k+1} = 2\beta_i^k$, caso contrario $\beta_i^{k+1} = \beta_i^k$. Como $\forall k \in \mathbb{N}$, $\|\mu^k\|_\infty \leq Q$, tenemos que si para $k_1 \in \mathbb{N}$, $\beta^{k_1} \geq Q$, entonces $\beta^k = \beta^{k_1}$ para todo $k \geq k_1$.

∎

Teorema 4.1.7 *(ver [3], Teorema 5.18)*
La sucesión $\{\mu^k\}$ generada por el método de punto proximal con distancia de Bregman (4.14) converge a una solución óptima del problema $(\bar{D})$.

El siguiente Teorema establece la factibilidad ergódica y la convergencia ergódica de una sucesión $\{\bar{x}^k\}$ de medias primal asociada a la sucesión $\{x^k\}$, de igual forma se establece la complementaridad de la sucesión $\{x^k\}$. Un resultado parecido presenta Castillo en ([3], Teorema 5.22) para las familias de penalidades AL1 y Al2, no obstante realiza la demostración para la penalidad del tipo AL2, dejando planteado el caso para AL1. La demostración que se presenta a continuación es para la penalidad θ definida en (4.1), la cual cumple con las caracteristícas de las penalidades de tipo AL1, excepto que no es estrictamente convexa.
Además se establece la acotación de la sucesión $\{x^k\}$, así como su convergencia a una solución óptima del problema (P) sin necesidad de considerar las dos condiciones adicionales que en el mencionado Teorema se imponen.

Teorema 4.1.8 *(Ver [3], Teorema 5.22)*
Suponga válidas las hipótesis A1 y A2 con respecto al problema (P). Consideremos las sucesiones $\{x^k\}, \{\mu^k\}, \{r^k\}, \{\beta^k\}$. generadas por el método de multiplicadores con shifts definidas en el Algoritmo 2.1.1 usando como penalidad la función θ. Considere además la sucesión de medias primal $\{\bar{x}^k\}$, con:

$$\bar{x}^k = \frac{\sum_{j=1}^k x^j/r^j}{\sum_{j=1}^k 1/r^j}.$$

Entonces:

i) Para $i = 1, \ldots, m$ se verifica $\lim_{k\to+\infty} \sup f_i(\bar{x}^k) \leq 0$. (factibilidad ergódica).

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

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

iv) Para $i = 1, \ldots, m,$ $\lim\limits_{k \to +\infty} f_i(x^k) \leq 0,$ $\{x^k\}$ *es acotada y cualquier punto límite de* $\{x^k\}$ *es solución óptima de (P).*

Demostración .

Consideremos la sucesión $\{\tilde{y}_i^k\}$ tal que $\theta'(\tilde{y}_i^k) = \frac{\mu_i^k}{\beta_i^k},$ para $i = 1, \ldots, m.$

i) Mostremos que para $i = 1, \ldots, m$; se verifica $\lim\limits_{k \to \infty} \sup f_i(\bar{x}^k) \leq 0.$
Por la Proposición 4.1.6 se tiene que la sucesión $\{\mu^k\}$ es acotada y existe un $k_1 \in \mathbb{N}$ tal que $\beta^k = \beta^{k_1}$, $\forall k \geq k_1$. Por lo cual

$$\tilde{y}_i^{k+1} = \tilde{y}_i^k + \frac{y_i^{k+1}}{r^k}, \quad \forall k \geq k_1. \tag{4.20}$$

Donde $\theta'(\tilde{y}_i^{k+1}) = \frac{\mu_i^{k+1}}{\beta_i^{k_1}},$ $\theta'(\tilde{y}_i^k) = \frac{\mu_i^k}{\beta_i^{k_1}},$ $y_i^{k+1} = f_i(x^{k+1})$ y $r^k \in (0, 1].$
Tomando la suma en ambos miembros de (4.20) obtenemos para $i = 1, \ldots, m$:

$$\tilde{y}_i^k = \tilde{y}_i^{k_1} + \sum_{j=k_1}^{k} \frac{y_i^j}{r^j}. \tag{4.21}$$

Como $f_i(\cdot)$ es convexa para $i = 1, \ldots, m$ tenemos:

$$f_i\left(\frac{\sum_{j=1}^k x^j/r^j}{\sum_{j=1}^k 1/r^j}\right) \leq \frac{\sum_{j=1}^k f_i(x^j)/r^j}{\sum_{j=1}^k 1/r^j} = \frac{\sum_{j=1}^{k_1-1} y_i^j/r^j}{\sum_{j=1}^k 1/r^j} + \frac{\sum_{j=k_1}^k y_i^j/r^j}{\sum_{j=1}^k 1/r^j}. \tag{4.22}$$

Por otro lado para cada $i = 1, \ldots, m$ se verifica $\theta'(\tilde{y}_i^k) = \frac{\mu_i^k}{\beta_i^k},$ $\forall k \in \mathbb{N},$ además por (4.3) tenemos que $0 \leq \theta'(x) \leq 1,$ $\forall x \in \mathbb{R}$, se deduce que $0 \leq \frac{\mu_i^k}{\beta_i^k} \leq 1,$ $\forall k \in \mathbb{N}.$
Luego, usando (4.5) se tiene:

$$-1 \leq \tilde{y}_i^k = (\theta')^{-1}(\frac{\mu_i^k}{\beta_i^k}) \leq 1,$$

por lo cual $\{\tilde{y}_i^k\}$ es acotada para cada $i = 1, \ldots, m.$
De (4.21) obtenemos:

$$\sum_{j=1}^{k} \frac{y_i^j}{r^j} = \sum_{j=1}^{k_1-1} \frac{y_i^j}{r^j} + \sum_{j=k_1}^{k} \frac{y_i^j}{r^j} \leq \sum_{j=1}^{k_1-1} \frac{y_i^j}{r^j} + 1 - \tilde{y}_i^{k_1}. \tag{4.23}$$

Dividiendo (4.23) entre $\sum\limits_{j=1}^{k} 1/r^j$, y usando (4.22) obtenemos:

$$f_i(\bar{x}_i^k) \leq \frac{\sum_{j=1}^{k_1} \frac{y_i^j}{r^j} + M - \tilde{y}_i^{k_1}}{\sum_{j=1}^{k} 1/r^j}.$$

Como $r^k \leq 1$ para $k \in \mathbb{N}$, se tiene que $\sum_{j=1}^{k} 1/r^j \to \infty$. De donde se deduce que:

$$\lim_{k\to\infty} \sup f_i(\bar{x}^k) \leq 0.$$

ii) La función φ es cóncava y por (4.19) $y^{k+1} \in \partial\varphi(\mu^{k+1})$, por lo tanto:

$$\sum_{i=1}^{k} (\mu_i^k - \mu_i^{k+1}) y_i^{k+1} \geq \varphi(\mu^k) - \varphi(\mu^{k+1}).$$

Luego

$$\sum_{i=1}^{k} \mu_i^{k+1} \left(\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \right) y_i^{k+1} \geq \varphi(\mu^k) - \varphi(\mu^{k+1}).$$

De donde:

$$\sum_{k=1}^{\infty} \sum_{i=1}^{m} \mu_i^{k+1} \left(\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \right) y_i^{k+1} \geq \varphi(\mu^0) - \lim_{k\to\infty} \varphi(\mu^k) > -\infty.$$

Sabemos por (4.20) que:

$$\tilde{y}_i^{k+1} - \tilde{y}_i^k = \frac{y_i^{k+1}}{r^k}, \quad \forall k \geq k_1.$$

Donde $r^k \in (0, 1]$. Como θ es convexa se tiene que la derivada $\mu_i = \theta'(\tilde{y}_i)$ es no decreciente. De esta forma si $y_i^{k+1} \geq 0$, tenemos que $\tilde{y}_i^{k+1} \geq \tilde{y}_i^k$, por lo cual $\mu_i^{k+1} \geq \mu_i^k$, de donde $\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \leq 0$.
Si $y_i^{k+1} < 0$, se deduce $\tilde{y}_i^{k+1} < \tilde{y}_i^k$, por tanto $\mu_i^{k+1} < \mu_i^k$, y así $\frac{\mu_i^k}{\mu_i^{k+1}} - 1 > 0$.
Se concluye que a partir de k_1 todos los términos de la suma anterior son no positivos y como tal suma está acotada inferiormente se puede deducir que:

$$\sum_{i=1}^{m} \mu_i^{k+1} \left(\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \right) y_i^{k+1} \to 0.$$

Como cada uno de los términos es de igual signo se deduce para i=1,...,m

$$\mu_i^{k+1}\left(\frac{\mu_i^k}{\mu_i^{k+1}}-1\right)y_i^{k+1}\to 0. \tag{4.24}$$

Por el Teorema 4.1.7 $\{\mu^k\}$ es convergente.
Si para $i=1,\ldots,m$ $\mu_i^k\to\mu^*>0$, entonces para un k suficientemente grande se tiene:

$$\frac{\mu_i^k}{\beta_i^{k_1}}=\theta'(\tilde{y}_i^k),$$

de donde $0\leq\frac{\mu_i^k}{\beta_i^{k_1}}\leq 1$, por lo que podemos calcular la inversa:

$$\lim_{k\to\infty}\tilde{y}_i^k=(\theta')^{-1}\left(\frac{\mu_i^*}{\beta_i^{k_1}}\right).$$

Por otro lado: $\frac{y_i^k}{r^k}=\tilde{y}_i^{k+1}-\tilde{y}_i^k$ y. Tomando límite tenemos $\frac{y_i^k}{r^k}\to 0$, y como $r^k\in(0,1]$, se deduce que $y_i^k\to 0$ y tenemos que $\lim_{k\to\infty}\mu_i^k y_i^k$.
Consideremos ahora el caso cuando $\mu_i^k\to 0$. Supongamos que para un $i=1,\ldots,m$ se cumple

$$|\mu_i^{k+1}y_i^{k+1}|\geq\epsilon>0,\quad\forall k\in\mathbb{N}. \tag{4.25}$$

Por (4.24) y por (4.25) se obtiene:

$$\frac{\mu_i^k}{\mu_i^{k+1}}\to 1. \tag{4.26}$$

Por (4.25) tenemos que $|y_i^{k+1}|\geq\frac{\epsilon}{\mu_i^{k+1}}$.
Además $\theta(x)\geq 0\quad\forall x\in\mathbb{R}$. Luego $\inf\limits_{t\in\mathbb{R}}\theta(t)=0$.
Podemos considerar dos casos:

a) Si $y_i^{k+1}>0$, usamos $\tilde{y}_i^{k+1}-\tilde{y}_i^k=\frac{y_i^{k+1}}{r^k}$; por la convexidad de θ y $\frac{\mu_i^k}{\beta_i^k}=\theta'(\tilde{y}_i^k)$, se tiene:

$$\begin{aligned}\theta(\tilde{y}_i^{k+1}) &\geq \theta(\tilde{y}_i^k)+\theta'(\tilde{y}_i^k)(\tilde{y}_i^{k+1}-\tilde{y}_i^k)\\ &= \theta(\tilde{y}_i^k)+\frac{\mu_i^k y_i^{k+1}}{\beta_i^k r^k}\\ &\geq \frac{\epsilon}{\beta_i^k r^k}\frac{\mu_i^k}{\mu_i^{k+1}}.\end{aligned}$$

Como $r^k \in (0,1]$, se tiene que $r^k < 2$ y por (4.26) se obtiene para un k suficientemente grande:

$$\theta(\tilde{y}_i^{k+1}) > \frac{\epsilon}{2\beta_i^{k_1}}.$$

luego $\tilde{y}_i^{k+1} > \theta^{-1}(\frac{\epsilon}{2\beta_i^{k_1}}) \equiv \gamma$, ya que θ es creciente.

Por tanto para k suficientemente grande $\frac{\mu_i^{k+1}}{\beta_i^{k_1}} = \theta'(\tilde{y}_i^{k+1}) > \theta'(\gamma) > 0$, lo cual contradice el hecho que $\mu_i^k \to 0$.

b) Si $y_i^{k+1} < 0$, por la convexidad de θ se deduce:

$$\begin{aligned}\theta(\tilde{y}_i^k) &\geq \theta(\tilde{y}_i^{k+1}) - \frac{\mu_i^k y_i^{k+1}}{\beta_i^{k_1} r^k} \\ &\geq \frac{\epsilon}{r^k} > \frac{\epsilon}{2\beta_i^{k_1}}.\end{aligned}$$

Luego $\tilde{y}_i^k > \theta^{-1}(\frac{\epsilon}{2\beta_i^{k_1}}) \equiv \gamma$, y así $\frac{\mu_i^k}{\beta_i^{k_1}} = \theta'(\tilde{y}_i^k) > \theta'(\gamma) > 0$, contradictorio con $\mu_i^k \to 0$.

Podemos concluir que $\lim\limits_{k\to\infty} \mu_i^k f(x_i^k) = 0$.

iii) Sabemos que x^k minimiza al lagrangeano $L(\cdot, \mu^k)$, luego usando la parte ii) tenemos:

$$\begin{aligned}\lim_{k\to+\infty} \varphi(\mu^k) &= \lim_{k\to+\infty} [f(x^k) + \textstyle\sum_{i=1}^m \mu_i^k f_i(x^k)] \\ &= \lim_{k\to+\infty} f(x^k).\end{aligned} \tag{4.27}$$

Por el Teorema de dualidad fuerte (Teorema 2.1.2), se deduce que:

$$\varphi(\mu^k) \leq \varphi(\mu^*) = f(x^*), \quad \forall k. \tag{4.28}$$

Por la definición de $\{\bar{x}^k\}$ y la convexidad de de f tenemos:

$$f(\bar{x}) = \frac{\sum_{j=1}^k f(x^j)/r^j}{\sum_{j=1}^k r^j}.$$

Por lo cual:

$$\lim_{k\to\infty} \sup f(\bar{x}^k) \leq \lim_{k\to\infty} \frac{\sum_{j=1}^k f(x^j)/r^j}{\sum_{j=1}^k r^j} = f(x^*), \tag{4.29}$$

ya que $f(x^k) \to f(x^*)$ y $r^k \in (0,1]$.
Dado $\epsilon > 0$, para k suficientemente grande de (4.29) y de i) tenemos:

$$\bar{x}^k \in \{x \in \mathbb{R}^n : \ f(x) \leq f(x^*) + \epsilon, \ f_i(x) \leq \epsilon, i = 1, \dots, m\}.$$

Usando ([3], Corolario 5.21) se tiene que este conjunto es compacto y por lo tanto $\{\bar{x}^k\}$ es acotada.
Dado $\widehat{x}$, un punto límite de $\{\bar{x}^k\}$, entonces $\bar{x}^k \xrightarrow{K} \widehat{x}$, con $K \in \mathbb{N}$. De esta forma $f(\widehat{x}) \leq f(x^*)$ y $\widehat{x}$ es factible, por lo cual $\widehat{x}$ es una solución óptima de (P).

iv) Sabemos que $\mu^k \to \mu^*$, consideramos $i = 1, \dots, m$. Si $\mu_i^k \to \mu_i^* > 0$, entonces por ii) se tiene $\lim_{k\to\infty} f_i(x^k) = 0$.
Si $\mu_i^k \to 0$ entonces $\tilde{y}_i^k \to -1$ y $\tilde{y}_i^{k+1} \to -1$.
Como:

$$\tilde{y}_i^{k+1} - \tilde{y}_i^k = \frac{y_i^{k+1}}{r^k}, \quad \forall k \geq k_1.$$

Tomando límite nos queda: $\frac{y_i^k}{r^k} \to 0$, por lo cual:

$$\lim_{k\to\infty} \sup f_i(x^{k+1}) \leq 0.$$

Por otro lado Finalmente de (4.27) y (4.28) nos queda:

$$\lim_{k\to\infty} f(x^k) = \lim_{k\to\infty} \varphi(\mu^k) \leq \varphi(\mu^*) = f(x^*), \tag{4.30}$$

luego:

$$x^k \in \{x \in \mathbb{R}^n : f(x) \leq f(x^*) + \epsilon, \ f_i(x) \leq \epsilon, i = 1, \dots, m\}.$$

Usando nuevamente ([3], Corolario 5.21) se tiene que este conjunto es compacto y por tanto $\{x^k\}$ es acotada.
Además por (4.30) se deduce que cada punto límite de $\{x^k\}$ es una solución óptima de (P).

$\blacksquare$

4-2 IMPLEMENTACIÓN NUMÉRICA

Finalmente presentamos en este trabajo una implementación numérica del método de mutiplicadores basados en shifts con la penalidad θ, ya estudiada. El propósito es realizar pruebas computacionales que muestren el desempeño del método considerado, en virtud de lo cual se realizó un programa computacional usando el software matemático Matlab, se incluye un programa principal, el cual usa varias subrutinas.
En esta sección mostramos el programa principal y posteriormente lo usaremos para dar una aproximación de la solución de algunos problemas de optimización convexa.

PROGRAMA PRINCIPAL:

```
% Método de multiplicadores con shifts usando una penalidad especifica.
format short
clear all
  global  objetivo epsibfgs epsigold epsilon n m restrito teta...
          restric delta ...
          kmax ifeval igradobjetivo ibfgslanau igolden ipenalty irestric
ifeval=0; igradobjetivo=0; ibfgslanau=0; igolden=0; ipenalty=0; irestric=0;
   problema=input ('Problema (P1): ','s');
if length(problema)==0 problema='P1'; end;
eval(problema);
xk=x0;
rk=r0;
betak=beta0;
muk=mu0;
ytildek1=0;
for i=1:m,
```

```
    ytildek(i)=ytilde(muk(i),betak(i));
end
if n== 2 disp(' x(1) x(2) muk(1) muk(2) fobjetivo fiagrangeana'),
fprintf(' %8.8f %10.8f %10.8f %10.8f %10.8f %10.8f \n',
xk(1), xk(2), muk(1), muk(2), feval('f',xk),
feval('fianau', xk, rk, betak, ytildek ) );
end
k=0;
goon=1;
while goon==1,
xk1=bfgslanau(xk,rk,betak,ytildek);
% pause
if norm(xk1-xk)<.0001 |(k==kmax), goon=0; end;
yk1=feval(restric,xk1);
irestric=irestric+1;
for i=1:m,    %Actualizacion del multiplicador
muk1(i)=betak(i)*tetakg(yk1(i)/rk(i)+ytildek(i));
end
rk=rk/2;
% Regla de parada con el multiplicador
for i=1:m,
 if     muk1(i)>betak(i)/2,
        betak1(i)=2*betak(i);
        ytildek1(i)=ytilde(muk1(i),betak1(i));
 else
        betak1(i)=betak(i);
        ytildek1(i)=ytildek(i)+yk1(i)/rk(i);

 end
end
```

```
betak=betak1;
xk=xk1;
ytildek=ytildek1;
if n==2,
plot(x0(1),x0(2),'*')
hold on
plot(xk1(1),xk1(2),'*')
hold on
fprintf(' %8.8f %10.8f %10.8f %10.8f %10.8f %10.8f %10.8f
%10.8f %10.8f \n', xk1(1),xk1(2),muk1(1),muk1(2),
feval('f',xk), feval('fianau',xk,rk,betak,ytildek), betak(i),betak(2),rk(i));
hold on
end;
k=k+1;
end
y=f(xk)
fprintf('Nro de evaluaciones de la funcion %8.0f\n', ifeval);
fprintf('Evaluaciones del Gradiente %8.0f\n', igradobjetivo);
fprintf('Nro de busquedaslineales %8.0f\n', igolden);
fprintf('Nro de iteraciones del BFGS %8.0f\n', ibfgslanau);
zoom on.
```

A continuación presentamos algunos problemas propuestos en [7] y [2] donde la función objetivo es una función convexa y las funciones asociadas a las restricciones también son convexas, estos problemas son resueltos usando el mencionado programa y presentamos como solución las salidas que arroja este:

Problema 1:(Ver [7], Problema 22)

$$\begin{array}{ll} \text{minimizar} & f(x) = 0.77(x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{sujeto a} & x_1^2 - x_2 \leq 0 \\ & x_1 + x_2 - 2 \leq 0. \end{array}$$

Solución:

$x(1)$	$x(2)$	$muk(1)$	$muk(2)$	fobjetivo	fiagrangeana
0.50000000	0.50000000	0.16666667	0.16666667	1.98250000	1.92868056
1.21180099	1.04280393	0.37949552	0.29396913	0.48020059	0.96242824
0.99685734	1.04203346	0.54401548	0.46016238	0.77661411	0.77592189
0.99918281	1.00491846	0.51780680	0.50725571	0.77128318	0.77038154
1.00108861	0.99959388	0.51227438	0.51271566	0.76832461	0.77011286
0.99945467	1.00012858	0.51344743	0.51150754	0.77084006	0.77005412
1.00026774	0.99991466	0.51381262	0.51067622	0.76958775	0.77002671
0.99987572	1.00004578	0.51484373	0.51148886	0.77019141	0.77001160
1.00006264	0.99999104	0.51319005	0.51333599	0.76990354	0.77000534
0.99997593	1.00000291	0.51730674	0.51479028	0.77003706	0.77000133

Tabla 4.1: Sucesiones x^k, μ^k, $f(x^k)$ y $L(x^k, \mu^k)$, Problema 1

Nro de evaluaciones de la función	507
Evaluaciones del Gradiente	30
Nro de busquedas lineales	21
Nro de iteraciones del BFGS	9

Tabla 4.2: Resumen Problema 1

Minimizador aproximado:

$$(x_1^*, x_2^*) = (0.99997593, 1.00000291)$$

Mínimo aproximado:

$$f(x_1^*, x_2^*) = 0.7700$$

Figura 4.4: Comportamiento de x^k, Problema 1.

Problema 2: (Ver [7], Problema 14)

$$\begin{array}{ll} \text{minimizar} & f(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{sujeto a} & 0,25x_1^2 + x_2 - 1,5 \leq 0 \\ & x_1 - 2x_2 + 1 \leq 0. \end{array}$$

Solución:

Minimizador aproximado:

$$(x_1^*, x_2^*) = (1.15832334, 1.07918620)$$

Mínimo aproximado:

$$f(x_1^*, x_2^*) = 0.7147.$$

x(1)	x(2)	muk(1)	muk(2)	fobjetivo	fiagrangeana
0.90000000	0.90000000	0.16666667	0.16666667	1.22000000	1.21138889
1.65289622	1.06979588	0.33040652	0.42331890	0.12535250	0.80289211
1.35801383	1.10047628	0.66624481	0.83703240	0.42224172	0.77693901
1.21976931	1.08573733	0.86938414	1.03021104	0.61611082	0.74930782
1.15498922	1.07452461	0.97734870	1.12525109	0.71959714	0.71804740
1.15940588	1.08124696	0.96458154	1.12147397	0.71319955	0.71616363
1.15786687	1.07814731	0.96908630	1.12328164	0.71529520	0.71539781
1.15850956	1.07964860	0.96652375	1.12309025	0.71445005	0.71503945
1.15821971	1.07891318	0.96786517	1.12296939	0.71482135	0.71486829
1.15835725	1.07927679	0.96715407	1.12314057	0.71464732	0.71478512
1.15829032	1.07909610	0.96748771	1.12309631	0.71473137	0.71474423
1.15832334	1.07918620	0.96732529	1.12311386	0.71469006	0.71472393

Tabla 4.3: Sucesiones x^k, μ^k, $f(x^k)$ y $L(x^k, \mu^k)$, Problema 2

Nro de evaluaciones de la función	765
Evaluaciones del Gradiente	41
Nro de busquedas lineales	30
Nro de iteraciones del BFGS	11

Tabla 4.4: Resumen Problema 2

Problema 3:(Ver [2],Ejercicio 4.7)

$$\begin{array}{ll} \text{minimizar} & f(x_1,x_2) = (x_1 - \frac{9}{4})^2 + (x_2 - 2)^2 \\ \text{sujeto a} & x_1^2 - x_2 - 1 \leq 0 \\ & x_1 + x_2 - 6 \leq 0. \end{array}$$

Solución:

Minimizador aproximado:

$$(x_1^*, x_2^*) = (1.77066217, 2.13534635)$$

x(1)	x(2)	muk(1)	muk(2)	fobjetivo	fiagrangeana
1.00000000	2.00000000	0.16666667	0.16666667	1.56250000	1.50694444
1.80948081	2.12107419	0.24323997	0.00000000	0.20871612	0.26942138
1.76002078	2.13910233	0.27838411	0.00000000	0.25942909	0.25132862
1.77495638	2.13486578	0.26816366	0.00000000	0.24385522	0.24901373
1.76876439	2.13552486	0.27138278	0.00000000	0.24995470	0.24844744
1.77162028	2.13524213	0.27056352	0.00000000	0.24713759	0.24824106
1.77023033	2.13540433	0.27071138	0.00000000	0.24851327	0.24814734
1.77091849	2.13531731	0.27040894	0.00000000	0.24782986	0.24810028
1.77057798	2.13535743	0.27082126	0.00000000	0.24816711	0.24807741
1.77074691	2.13534043	0.27064764	0.00000000	0.24800056	0.24806649
1.77066217	2.13534635	0.27070854	0.00000000	0.24808339	0.24806113

Tabla 4.5: Sucesiones x^k, μ^k, $f(x^k)$ y $L(x^k, \mu^k)$, Problema 3

Figura 4.5: Comportamiento de x^k, Problema 3.

Mínimo aproximado:

$$f(x_1^*, x_2^*) = 0.2481.$$

Nro de evaluaciones de la función	590
Evaluaciones del Gradiente	31
Nro de busquedas lineales	21
Nro de iteraciones del BFGS	10

Tabla 4.6: Resumen del Problema 3

Problema 4:(Ver [2], Ejemplo 6.4.2)

$$\begin{array}{ll} \text{minimizar} & f(x_1, x_2) = x_1^2 + x_2^2 \\ \text{sujeto a} & -x_1 - x_2 + 3.5 \leq 0 \\ & x_1 + 2x_2 - 6 \leq 0. \end{array}$$

Solución:

x(1)	x(2)	muk(1)	muk(2)	fobjetivo	fiagrangeana
3.00000000	3.00000000	0.16666667	0.16666667	18.00000000	20.27777778
0.50173197	0.49802584	1.00000000	0.00000000	0.49976471	5.25024908
1.00000166	1.00000168	2.00000000	0.00000000	2.00000668	7.74999332
1.66666678	1.66666678	3.33333153	0.00000000	5.55555631	6.54860979
1.74741762	1.74741763	3.49860354	0.00000000	6.10693671	6.12671330
1.75121446	1.75121447	3.50842394	0.00000000	6.13350421	6.12573765
1.74943022	1.74943021	3.49883713	0.00000000	6.12101216	6.12533177
1.75028154	1.75028154	3.50055489	0.00000000	6.12697092	6.12516218
1.74986006	1.74986006	3.49970733	0.00000000	6.12402045	6.12508017
1.75006976	1.75006976	3.50014619	0.00000000	6.12548830	6.12503985
1.74996518	1.74996518	3.49992436	0.00000000	6.12475623	6.12501987
1.75001701	1.75001701	3.50321000	0.00000000	6.12511908	6.12500937

Tabla 4.7: Sucesiones x^k, μ^k, $f(x^k)$ y $L(x^k, \mu^k)$, Problema 4

Minimizador aproximado:

$$(x_1^*, x_2^*) = (1.75001701, 1.75001701)$$

Mínimo aproximado:

$$f(x_1^*, x_2^*) = 6.1251$$

Nro de evaluaciones de la función	651
Evaluaciones del Gradiente	33
Nro de busquedas lineales	22
Nro de iteraciones del BFGS	11

Tabla 4.8: Resumen del Problema 4

Problema 5:(Ver [2], Ejercicio 4.24)

$$\begin{array}{ll} \text{minimizar} & f(x_1,x_2) = x_1^2 - x_1x_2 + 2x_2^2 - 4x_1 - 5x_2 \\ \text{sujeto a} & x_1 + 2x_2 - 6 \leq 0 \\ & x_1 - 2 \leq 0. \end{array}$$

Solución:

x(1)	x(2)	muk(1)	muk(2)	fobjetivo	flagrangeana
1.50000000	1.00000000	0.16666667	0.16666667	-8.25000000	-8.30555556
2.55559122	1.77774626	0.22220853	0.44446228	-10.80247890	-10.24843467
2.31883211	1.76569871	0.12797993	1.00000000	-10.58581463	-10.07315041
2.13047039	1.78260177	0.00000000	1.52188156	-10.33842847	-10.00368781
2.01265960	1.75353026	0.00000000	1.72443517	-10.14701380	-10.12005465
1.99477615	1.74873169	0.00000000	1.75982559	-10.11583438	-10.12328098
2.00238897	1.74872940	0.00000000	1.74555645	-10.12916873	-10.12426813
1.99889079	1.74878618	0.00000000	1.75647127	-10.12305605	-10.12468938
2.00052835	1.75011815	0.00000000	1.74974997	-10.12592437	-10.12485696
1.99973787	1.74994003	0.00000000	1.75079903	-10.12454122	-10.12492979
2.00013631	1.75026480	0.00000000	1.75617648	-10.12523843	-10.12496098
1.99992890	1.74998300	0.00000000	1.75014063	-10.12487556	-10.12497930

Tabla 4.9: Sucesiones x^k, μ^k, $f(x^k)$ y $L(x^k, \mu^k)$, Problema 5

Minimizador aproximado:

$$(x_1^*, x_2^*) = (1.99992890, 1.74998300)$$

Mínimo aproximado:

$$f(x_1^*, x_2^*) = -10.1249.$$

Nro de evaluaciones de la función	787
Evaluaciones del Gradiente	40
Nro de busquedas lineales	29
Nro de iteraciones del BFGS	11

Tabla 4.10: Resumen del Problema 5

Los problemas que se han resuelto han sido presentados por Hock y Schittkowski en [7] y Bazaraa y Shetty en [2], al aplicar el programa se han obtenido soluciones muy cercanas a las expresadas en [7] y en [2], por lo cual podemos deducir que el método de multiplicadores basados en shifts con la penalidad no coerciva θ tiene buen desempeño computacional, en el sentido que encuentra las soluciones óptimas. Por otro lado se puede observar que el valor mínimo del problema primal coincide, con cierta tolerancia, con el valor óptimo de la función lagrangeano aumentado.

Otro aspecto importante observado es que el número de evaluaciones de la función objetivo y de los gradientes no es muy elevado, de igual forma los problemas convergen, con la tolerancia fijada, en un número razonable de iteraciones.

Los problemas que se han analizados son problemas que no revisten mayor complejidad, además son problemas de dos variables y dos restricciones puesto que el objetivo en este caso es mostrar la eficiencia del método para encontrar los óptimos, no obstante queda abierta la posibilidad de hacer análisis para problemas más complejos y de mayor número de variables y restricciones en próximas investigaciones.

CONCLUSIONES

El método de multiplicadores inexactos presentados por Kiwiel en [9] difiere del método de multiplicadores basados en shitfs presentados por Castillo en [3] en que la función de penalidad que emplea Kiwiel en su método es coerciva, en cambio Castillo presenta un algoritmo para penalidades no coercivas de tipo AL1 y otro para penalidades coercivas de tipo AL2, en este sentido se hace posible el uso de la penalidad θ, la cual no es coerciva en el método de multiplicadores basados en shifts.

La penalidad θ que se consideró en esta investigación es propia, cerrada, convexa, continua, esencialmente suave, no coerciva y cumple las propiedades de AL1, excepto la convexidad estricta. Además la conjugada de θ es la función $\theta^*(y) = y(y-1)$, con $\mathcal{D}_{\theta^*} = [0,1]$. Como la función de penalidad ψ usada por Kiwiel es coerciva y verifica lo siguiente $\mathbb{R}_> \subset \operatorname{im} \nabla\psi \subset \mathbb{R}_+$, entonces θ no es considerada por Kiwiel en su método, ya que θ es no coerciva y además $\operatorname{im} \nabla\theta = [0,1]$. Por otro lado θ no es estrictamente convexa, por lo cual no pertenece a la familia AL1, de donde tampoco ha sido considerada por Castillo, no obstante esta condición de convexidad estricta no restringe el uso de esta penalidad en el método de multiplicadores basados en shifts.

El método de multiplicadores basados en shitfs con la penalidad θ es eficiente desde el punto de vista teórico, ya que demostramos que la sucesión $\{\mu^k\}$ es convergente, además se estableció que la sucesión $\{x^k\}$ es acotada y converge a una solución óptima. Este resultado se obtuvo sin necesidad de usar dos condiciones que usó Castillo para probar un resultado similar en [3].

Asimismo se puede decir que el método de multiplicadores basados en shifts con la penalidad no coerciva θ es eficiente desde el punto de vista computacional, ya que se realizó un programa en Matlab que logró hallar una buena aproximación a las soluciones óptimas de algunos problemas en un número razonable de iteraciones.

Bibliografía

[1] Auslender, A., Cominetti, R. and Haddou, M. *Asymptotic Analysis for penalty and barrier methods in convex and linear programming.* Mathematics of Operations Research, vol 22, Nro. 1 (1997), pp. 43-62.

[2] Bazaraa, M. and Shetty, C. *Nonlinear Programming, Theory and Algorithms*: John Wiley & Sons. New York. 1779.

[3] Castillo, R. *Penalidades generalizadas y métodos de multiplicadores para programación no-lineal.*DSc. COPPE. UFRJ. Brasil. 1998.

[4] Chen, G. and Teboulle, M. *Convergence analysis of a proximal-like minimization algorithm using Bregman functions* Siam J. Optimization, 3, 1993. pp 538-543.

[5] Hamala, M. Quasibarrier Methods for convex programming. Mathematical Programming,20, 1985.

[6] Hiriart-Urruty, J. and Lemaréchal, C. *Convex Analysis and Minimization Algorithms I.* Springer-Verlag. Berlin. 1993.

[7] Hock, W. Schittkowski, K. *Test Examples for Nonlinear Programming Codes.* Springer-Verlag.1980.

[8] Iusem, A. *Metodos de ponto proximal.* 20 Coloquio Brasileiro de Matematica. Impa.1996.

[9] Kiwiel, K. *Proximal minimization methods with generalized Bregman functions.* SIAM J. on Optimization, 35, 1997. pp 1142-1168.

[10] Kiwiel, K. *Free-steering relaxation methods for problems with strictly convex costs and linear constraints*. Mathematics of operations research, 22, 1997. pp 326-346.

[11] Lara, H. y Castillo, R. *Optimización*: Universidad Centroccidental Lisandro Alvarado, Asociación Matemática Venezolana. II Talleres de Formación Matemática. 2001.

[12] Luenberguer, D. *Linear Programming and Nolinear*. Wesley & Son. 1980.

[13] Nash, S. and Sofer, A. *Linear and Nonlinear Programming*: McGraw-Hill. New York. 1996.

[14] Ortega, J. y González, J. *Cálculo V, Análisis Numérico*. Universidad Nacional Abierta. Caracas. 1984.

[15] Rockafeller, R. *Convex Analysis*. Princenton University Press. Princenton. N. J. 1970.

[16] Rockafeller, R. *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*. Mathematics of operations research, 1, 1976. pp 97-114.

[17] Vandenberghe, L. and Boyd, S. *Semidefinite Programming*: SIAM Review. 1996.

[18] Wegner, E. *Método de lagrangeano aumentado usando penalidades com conjugadas nao definidas ou nao coercivas na frontera*. Universidede Federal de Santa Catarina. 1998.

[19] Zang, I. *A Smoothing-out Technique for min-max Optimization*. Mathematical Programming, 19, 1980. pp 61-77.