

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”
UCLA

UN MÉTODO DE MULTILPLICADORES QUE CONDUCE A
UNA CUASIDISTANCIA HOMOGÉNEA DE TERCER ORDEN

LIC. JORGE CAMPOS

BARQUISIMETO, VENEZUELA

2005

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”

MAESTRÍA EN CIENCIAS

UCLA

UN MÉTODO DE MULTILPLICADORES QUE CONDUCE A
UNA CUASIDISTANCIA HOMOGÉNEA DE TERCER ORDEN

TRABAJO DE GRADO PRESENTADO COMO REQUISITO PARCIAL

PARA OPTAR AL TÍTULO DE MAGISTER SCIENTIARUM

MENCIÓN MATEMÁTICA

AUTOR: LIC. JORGE CAMPOS

TUTOR: DR. RÓMULO CASTILLO

BARQUISIMETO, VENEZUELA

2005

*A Dios, A Dorka, A mi Familia
Y en especial A mis Padres*

Agradecimientos

A Dios Omnisapiente, que sin su iluminación y guía no podría haber alcanzado esta y otras tantas metas en mi vida espiritual, personal y profesional.

A mis padres, que por medio de ellos se me otorgó el don de la vida. Mamá, eres ejemplo de humildad y paciencia. Papá, aunque estés ausente, tus sabias palabras y tu buen ejemplo siempre estarán vivos en mí.

A mi amada Dorka, que sin su amor y bondad hacia a mí el camino fuese sido más largo y tortuoso. Te amo mucho mi corazón.

A mi muy querida hermana Juanita, siempre tendrás un lugar especial en mi vida.

A mis hermanos y hermanas que siempre me han apoyado en cada una de las etapas de mi vida académica.

A mi amigo y tutor Rómulo Castillo, sus conocimientos y experiencia fueron invaluable en la realización del presente trabajo.

A mis amigos y colegas Eibar Hernández y Malón Mendoza, han sido unos faros de sabiduría en mi vida personal y laboral.

A mi amigo y colega Eduardo Villegas, “el morocho”, por su amistad incondicional.

A mi amigo y colega Abelardo Monsalve por su ayuda en la transcripción de estas líneas.

A todos mis compañeros de clase y mis colegas, que de una u otra forma me han ayudado en mis estudios.

A todos los profesores que me impartieron clase y me formaron intelectualmente.

A la UCLA donde recibí mi formación profesional.

A todos aquellos que que no nombro pero que han sido copartícipes en este trabajo, mil perdones por la omisión.

A todos mis más sinceras gracias.

Resumen

En el presente trabajo se considera el problema de programación convexa y en particular algunos métodos de resolución como los propuestos en [3] y en [9], los cuales están basados en métodos de multiplicadores, también conocidos como métodos de lagrangeano aumentado. En esta metodología se minimiza iterativamente una función auxiliar, que se conoce como función lagrangeano aumentado, e involucra una función de penalidad comunmente conocida. Se genera además, vía Dualidad de Fenchel, una cuasidistancia en el espacio dual desconocida hasta ahora, la cual es analizada. Presentamos un método de multiplicadores totalmente novedoso en el sentido que involucra potencias cúbicas de los multiplicadores en la función lagrangeano, lo cual no había sido considerado hasta la fecha, además presentamos un algoritmo que genera un par de sucesiones de las cuales enunciamos y demostramos resultados concernientes a su convergencia.

Finalmente, aún cuando nuestro objetivo principal es el estudio teórico y no el estudio numérico del método propuesto, mostramos un programa en MATLAB que fué implementado a fin de observar la eficacia computacional del mismo. Cabe destacar que los resultados obtenidos en este particular son bastante alentadores, basados en los seis ejemplos presentados al final de este trabajo, se pudo observar un buen desempeño del método.

Índice general

Agradecimientos	II
Resumen	III
Índice general	IV
Introducción	1
1. Conceptos Preliminares	6
1.1. Funciones Convexas	6
1.2. Dualidad	11
2. Métodos de Multiplicadores y Punto Proximal	14
2.1. Métodos de Punto Proximal	15
2.1.1. El Método de Punto Proximal Clásico	15
2.1.2. Método de Punto Proximal con Operadores Monótonos Maxi- males	15
2.1.3. Método de Punto Proximal con Cuasidistancias	19
2.2. Métodos de Lagrangeano Aumentado	22
2.2.1. El Lagrangeano Clásico	23
2.2.2. Lagrangeano Aumentado	24
3. Método Propuesto	27
3.1. Desarrollo del Método	29
3.2. Análisis de Convergencia	36
3.3. Implementación del Método	44
Conclusiones	59
Bibliografía	60

Introducción

En ingeniería, así como también en economía, es bastante común que se presenten problemas en los que es necesario minimizar costos de producción o maximizar ganancias, en general, las variables involucradas en el problema están sujetas a una o varias restricciones, las cuales varían con cada problema. Al formular el modelo matemático del problema en cuestión, nos damos cuenta que con mucha frecuencia aparecen funciones no lineales, estos problemas son conocidos como *problemas de programación no lineal*. Cuando las funciones involucradas en el problema son funciones convexas y continuamente diferenciables, el problema se convierte en un *problema de programación no lineal convexo diferenciable con restricciones*, digamos

$$(P) \quad \begin{array}{l} \text{minimizar } f_0(x) \\ \text{sujeto a } f_i(x) \leq 0, \quad i \in \{1, \dots, m\}. \end{array}$$

donde f_0, f_1, \dots, f_m son funciones convexas de clase C^1 en \mathbb{R}^n .

Los métodos de lagrangeano aumentado, llamados también métodos de multiplicadores, son los que generalmente se usan para resolver este tipo de problemas. Su buen desempeño computacional y su estrecha relación con los métodos de punto proximal, que se usan para resolver el problema dual de (P) , es la razón por la cual son usados tan frecuentemente en su resolución.

El *problema dual* de (P) , el cual es un problema cóncavo, viene dado por:

$$(D) \quad \sup\{d(\mu) : \mu \in \mathbb{R}_+^m\}$$

donde

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

es la *función dual objetivo* de (P) y

$$l : \mathbb{R}^n \times \mathbb{R}_{++}^m \rightarrow \mathbb{R}$$

dada por

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

es la *función lagrangeano clásico* para (P) .

En [9] Polyak y Teboulle presentan una metodología en métodos de multiplicadores que denominan “Nonlinear Rescaling” (NR), la cual consiste, básicamente, en transformar el problema inicial en uno equivalente cuya función lagrangeano $L_r(x, \mu) : \mathbb{R}^n \times \mathbb{R}_{++}^m$ viene dada por

$$L_r(x, \mu) = f_0(x) + r \sum_{j=1}^m \mu_j \theta \left(\frac{f_j(x)}{r} \right) \quad (0.0.1)$$

donde r es un número real positivo y $\theta : \mathbb{R} \rightarrow \mathbb{R}$ es una función de penalidad la cual definiremos en el capítulo 2. La función dada por (0.0.1) es la función objetivo del método tipo lagrangeano aumentado considerado por ellos. La cuasidistancia asociada a esta función vía conjugacidad, es conocida actualmente como φ -divergencia. Polyak y Teboulle presentan, en primer lugar, un método de multiplicadores para resolver (P) , generando un método de punto proximal que usan para resolver (D) . La “sucesión primal” generada por el método converge de manera “ergódica” a una solución óptima de (P) y la sucesión de “puntos duales” converge globalmente a una solución óptima de (D) .

Ben-Tal y Zibulevsky en [3] presentan un método de multiplicadores que llaman “The Penalty/Barrier Multiplier Method” (PBM) y usan para resolver (P) . Al considerar (D) , se encuentran con un método de punto proximal cuyo núcleo es una cuasidistancia que involucra una función sublineal. Al hacer el estudio de convergencia obtienen que cualquier punto límite de la “sucesión primal” converge a una solución óptima de (P) y cualquier punto límite de la “sucesión dual” converge a una solución óptima de (D) . La función lagrangeano aumentado $L(x, \mu, p) : \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ usada

por ellos viene dada por:

$$L(x, \mu, p) = f_0(x) + \sum_{j=1}^m \mu_j p_j \theta \left(\frac{f_i(x)}{p_j} \right) \quad (0.0.2)$$

donde θ es como en (0.0.1). Es de hacer notar que al tomar $p = r\mu$ en (0.0.2), con r un número real positivo, se genera, en el núcleo del método de punto proximal, una cuasidistancia homogénea de segundo orden.

Auslender, Teboulle y Ben-Tiba en [1], comienzan generando un algoritmo que llaman “Basic Iterations Scheme” (BIS), que es una versión aproximada del método de punto proximal, el cual usa una cuasidistancia homogénea de segundo orden y es usado para resolver (D) . Se estudian dos clases de algoritmos basados en BIS, estos son “Interior Proximal Method” (IPM) y “Regularized Interior Proximal Method” (RIPM). Al considerar (P) vía dualidad, utilizan un método de multiplicadores inexactos. En el análisis de convergencia se obtienen dos resultados para los dos algoritmos antes mencionados. La correspondiente función lagrangeano aumentado $L_r(x, \mu) : \mathbb{R}^n \times \mathbb{R}_{++}^m$ que se obtiene es dada por:

$$L_r(x, \mu) = f_0(x) + r \sum_{j=1}^m (\mu_j)^2 \theta \left(\frac{f_i(x)}{r\mu_j} \right) \quad (0.0.3)$$

donde r y θ son como en (0.0.1).

Al observar las funciones lagrangeano aumentado dadas por (0.0.1), (0.0.2) y (0.0.3), surge la interrogante en el sentido de analizar si es o no posible considerar potencias cúbicas de los multiplicadores que aparecen en las funciones objetivos y qué consideraciones y resultados adicionales se podrían obtener.

En esta dirección, hemos formulado nuestro principal objetivo, el cual consiste en obtener un método de multiplicadores alternativo, en este caso utilizando potencias cúbicas en los multiplicadores, que nos permita resolver (P) usando la función lagrangeano aumentado $L_r(x, \mu) : \mathbb{R}^n \times \mathbb{R}_{++}^m$ dada por:

$$L_r(x, \mu) = f_0(x) + r \sum_{j=1}^m (\mu_j)^3 \theta \left(\frac{f_i(x)}{r(\mu_j)^2} \right) \quad (0.0.4)$$

con r y θ como en (0.0.1).

Dicho esto, en el capítulo 1 haremos una síntesis de toda la teoría que usaremos en los capítulos subsecuentes, comenzando por el concepto de conjunto afín, pasando por funciones convexas, hasta llegar al teorema de dualidad de Fenchel.

En el capítulo 2 hablaremos del método de punto proximal clásico, referiremos algunos detalles de este método usando operadores monótonos maximales y hablaremos del método en su versión con cuasidistancias. Este método es una poderosa herramienta teórica, más no computacional, para los problemas de programación no lineal. Nuestra intención es aplicar este método para resolver (D) . Seguidamente haremos una síntesis de los métodos de lagrangeano aumentado. En la actualidad estos están siendo estudiados por muchos investigadores, quienes han obtenido grandes avances teóricos y computacionales en relación a estos. Los métodos de lagrangeano aumentado guardan una interesante relación con los métodos de punto proximal vía dualidad de Fenchel, lo cual es una gran ventaja, pues estos últimos, como ya dijimos antes, son una fuerte herramienta teórica. Uno de los problemas de los métodos de lagrangeano aumentado, es la dificultad que presentan a la hora de probar la convergencia de la “sucesión primal”.

Finalmente, en el capítulo 3, que es nuestro trabajo en sí, atacaremos (P) usando un método de multiplicadores con la función (0.0.4) que conduce, vía dualidad de Fenchel, a un método de punto proximal distinto a los considerados anteriormente en [9], [3] y [1], por tener un núcleo que es una cuasidistancia homogénea de tercer orden. Al igual que los métodos ya conocidos, este genera dos sucesiones, una sucesión de “puntos primales” y otra de “puntos duales”, mediante la aplicación de un algoritmo iterativo. Enunciaremos y demostraremos resultados de convergencia para estas dos sucesiones. El objetivo principal de este capítulo, y el de nuestro trabajo en general, es el estudio teórico del método propuesto, sin embargo, también incluimos

una implementación en MATLAB del mismo, con algunos ejemplos seleccionados adecuadamente, para estudiar su eficacia computacional.

Capítulo 1

Conceptos Preliminares

La idea de este capítulo es presentar gran parte de la terminología, junto con algunos teoremas y definiciones, que vamos a usar en el presente trabajo. El lector puede remitirse a la bibliografía, especialmente [10], para una prueba de los teoremas de este capítulo.

1.1. Funciones Convexas

Definición 1.1. Un conjunto $M \subset \mathbb{R}^n$ es llamado afín si

$$\forall x, y \in M, \forall \lambda \in \mathbb{R} : (1 - \lambda)x + \lambda y \in C.$$

Es claro que la intersección arbitraria de conjuntos afines es un conjunto afín.

Definición 1.2. Sean $S \subset \mathbb{R}^n$ cualquiera y $\mathcal{A} = \{M \subset \mathbb{R}^n : S \subset M \text{ y } M \text{ es afín}\}$.

Definiremos la cápsula afín de S como el conjunto $\text{aff}S = \bigcap_{M \in \mathcal{A}} M$.

La observación hecha inmediatamente antes de esta definición, nos permite garantizar que la cápsula afín de un conjunto $S \subset \mathbb{R}^n$, es siempre un conjunto afín, de hecho, $\text{aff}S$ es el conjunto afín más pequeño que contiene a S .

Definición 1.3. Un conjunto $C \subset \mathbb{R}^n$ es llamado convexo si

$$\forall x, y \in C, \forall \lambda \in [0, 1] : (1 - \lambda)x + \lambda y \in C.$$

Los conjuntos afines son ejemplos de conjuntos convexos. Otros ejemplos de conjuntos convexos, que no son afines, son las bolas (cerradas o abiertas) en \mathbb{R}^n , así como

también los paralelepípedos, los rectángulos, las pirámides y los triángulos. No son conjuntos convexos, la región en \mathbb{R}^3 entre dos esferas concéntricas, la región en \mathbb{R}^2 bajo la curva $y = \sin x$.

Teorema 1.1 (ver [10], Teorema 2.1). *La intersección arbitraria de conjuntos convexos es un conjunto convexo.*

Definición 1.4. Sean $S \subset \mathbb{R}^n$ cualquiera y

$$\mathcal{C} = \{M \subset \mathbb{R}^n : S \subset M \text{ y } M \text{ es convexo}\}.$$

Definiremos la cápsula convexa de S como el conjunto $\text{conv}S = \bigcap_{M \in \mathcal{C}} M$.

Según el teorema anterior, $\text{conv}S$ es un conjunto convexo, más aún, es el conjunto convexo más pequeño que contiene a S .

Definición 1.5. Sea f una función a valores reales extendidos con dominio $S \subset \mathbb{R}^n$. El conjunto

$$\text{epi}f = \{(x, \alpha) : x \in S, \alpha \in \mathbb{R}, \alpha \geq f(x)\}$$

es llamado el epígrafo de f .

Definición 1.6. Sea f una función definida sobre un conjunto $S \subset \mathbb{R}^n$. Diremos que f es convexa sobre S si $\text{epi}f$ es un subconjunto convexo de \mathbb{R}^{n+1} . Diremos que f es cóncava sobre S si $-f$ es convexa sobre S .

Definición 1.7. El dominio efectivo de una función convexa f definida sobre un conjunto $S \subset \mathbb{R}^n$, el cual denotaremos por $\text{dom}f$, está dado por

$$\text{dom}f = \{x \in S : f(x) < +\infty\}$$

Obviamente, la convexidad de f es equivalente a que la restricción de f sobre $\text{dom}f$ lo sea. Realmente el interés se centra sobre esta restricción, y S mismo tiene un rol secundario en esto.

Como lo afirma Rockafellar en [10], uno podría limitar su atención a las funciones las cuales nunca son $+\infty$, así que S siempre coincidiría con $\text{dom}f$. O uno podría limitar su atención a las funciones dadas sobre todo \mathbb{R}^n , dado que una función convexa f sobre S puede ser extendida a una función convexa sobre todo \mathbb{R}^n tomando $f(x) = +\infty$ para $x \notin S$.

El segundo enfoque es el que adoptaremos. Así que de ahora en adelante por una función convexa entenderemos una función convexa a valores reales extendidos la cual está definida en todo \mathbb{R}^n , a menos que se especifique otra cosa. Este enfoque tiene la ventaja que las molestias técnicas del dominio efectivo pueden ser suprimidas casi por completo. Por ejemplo, cuando una función convexa f es construida de acuerdo a ciertas fórmulas, las mismas fórmulas especifican el dominio efectivo de f de manera implícita, pues ésta especifica donde $f(x)$ es o no $+\infty$. En otras palabras, uno siempre tendría descrito el dominio efectivo de f explícitamente antes que los valores de f o de su dominio puedan darse.

Existen varias caracterizaciones de funciones convexas que en la práctica son más útiles que la definición en si, algunas de ellas son.

Teorema 1.2 (ver [10], Teorema 4.1). *Sea $C \subset \mathbb{R}^n$ un conjunto convexo. Una función $f : C \rightarrow (-\infty, +\infty]$ es convexa sobre C si y sólo si*

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

para todo $0 \leq \lambda \leq 1$ y para cualesquiera $x, y \in C$.

En muchos textos esta caraterización se usa como la definición de función convexa. Si la desigualdad en el teorema anterior es estricta, la función f es llamada *estrictamente convexa*.

Teorema 1.3 (ver [10], Teorema 4.2). *Sea f una función a valores reales extendidos definida sobre \mathbb{R}^n . Entonces f es convexa si y sólo si*

$$\forall \lambda \in (0, 1) : f((1 - \lambda)x + \lambda y) < (1 - \lambda)\alpha + \lambda\beta$$

siempre que $f(x) < \alpha$ y $f(y) < \beta$.

Teorema 1.4 (Desigualdad de Jensen)(ver [10], Teorema 4.3). *Sea f una función de \mathbb{R}^n en $(-\infty, +\infty]$. Entonces f es convexa si y sólo si*

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

siempre que $\lambda_1, \dots, \lambda_m \geq 0$ y $\sum_{i=1}^m \lambda_i = 1$.

Las funciones cóncavas satisfacen resultados análogos bajo hipótesis similares, salvo que las desigualdades serán contrarias. Otros resultados interesantes sobre la convexidad de funciones son.

Teorema 1.5 (ver [10], Teorema 4.4). *Una función f a valores reales dos veces diferenciable definida en un intervalo (a, b) es convexa si y sólo si f'' es no negativa en (a, b) .*

Teorema 1.6 (ver [10], Teorema 4.5). *Sea f una función a valores reales dos veces diferenciable sobre un conjunto convexo abierto $C \subset \mathbb{R}^n$. Entonces f es convexa si y sólo si su matriz Hessiana*

$$H(c) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(c) \right)_{n \times n}$$

es semidefinida positiva para todo $c \in C$.

Para funciones estrictamente convexas, se obtienen resultados similares, basta cambiar no negativa por positiva en el Teorema 1.5 y semidefinida por definida en el Teorema 1.6. Además, los análogos de los teoremas anteriores para funciones cóncavas,

se obtienen al cambiar no negativa por no positiva en el Teorema 1.5 y semidefinida positiva por semidefinida negativa en el Teorema 1.6.

Definición 1.8. Una función convexa f es llamada propia si su epígrafo es no vacío y no contiene rectas verticales, es decir, si $f(x) < +\infty$ para al menos un valor de x y $f(x) > -\infty$ para todo valor de x . Una función convexa que no es propia es llamada impropia.

En la siguiente definición $B[x, \varepsilon] = \{y : \|y - x\| \leq \varepsilon\}$.

Definición 1.9. Sea $C \subset \mathbb{R}^n$ un conjunto convexo. Definiremos el relativo interior de C al conjunto $\text{ri}C$ formado por todos los puntos $x \in \text{aff}C$ tales que

$$\exists \varepsilon > 0 : B[x, \varepsilon] \cap \text{aff}C \subset C.$$

El relativo interior de un conjunto convexo es también un conjunto convexo. El siguiente teorema es muchas veces más fácil de usar que la definición anterior.

Teorema 1.7 (ver [10], Corolario 6.4.1). Sea $C \subset \mathbb{R}^n$ un conjunto convexo no vacío. Entonces $x \in \text{ri}C$ si y sólo si para todo $y \in \mathbb{R}^n$ existe $\varepsilon > 0$ tal que $x + \varepsilon y \in C$.

Definición 1.10. Una función a valores reales extendidos f definida sobre un conjunto $S \subset \mathbb{R}^n$ es llamada semicontinua inferiormente en un punto $x \in S$ si

$$f(x) = \liminf_{y \rightarrow x} f(y) = \lim_{\varepsilon \rightarrow 0^+} \left(\inf_{|y-x| \leq \varepsilon} f(y) \right).$$

La *semicontinuidad superior* se define tomando \limsup en lugar de \liminf . Una función f se dice semicontinua inferiormente (o superiormente) en S si es semicontinua inferiormente (o superiormente) en todo punto $x \in S$. Es claro que f es semicontinua inferior y superiormente en x si y sólo si es continua en x .

Definición 1.11. Una función $f : \mathbb{R}^n \rightarrow (0, +\infty]$ es cerrada si $\text{epi}f$ es un conjunto cerrado en \mathbb{R}^{n+1} .

La importancia de la semicontinuidad inferior, en el estudio de la convexidad de funciones, es puesta de manifiesto en el siguiente teorema.

Teorema 1.8 (ver [10], Teorema 7.1). *Sea $f : \mathbb{R}^n \rightarrow [-\infty, +\infty]$. Entonces las siguientes condiciones son equivalentes*

1. f es semicontinua inferiormente en \mathbb{R}^n ;
2. $\{x : f(x) \leq \alpha\}$ es cerrado para todo $\alpha \in \mathbb{R}$;
3. f es cerrada.

1.2. Dualidad

El estudio de la dualidad en los problemas de programación convexa es de vital importancia. A continuación presentaremos algunos resultados, relativos a la dualidad, que nos serán útiles en el último capítulo.

En lo que resta del trabajo, si $x = (x_1, \dots, x_n)^T$ y $y = (y_1, \dots, y_n)^T$ son vectores en \mathbb{R}^n , entonces $xy = \sum_{i=1}^n x_i y_i$ denota el producto interno o producto escalar de x con y .

Definición 1.12. *Dada una función convexa $f : \mathbb{R}^n \rightarrow (0, +\infty]$, definiremos la conjugada de f como la función $f^* : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ dada por*

$$f^*(s) = \sup_x \{xs - f(x)\}.$$

Esta definición nos será muy útil en el capítulo 3, y es una pieza clave en el estudio del problema que nos proponemos presentar en este trabajo.

Se puede probar que f^* es una función convexa y cerrada, además f^* es propia si y sólo si f lo es. Cuando f es cóncava, la conjugada se define de manera análoga sólo que en lugar de tomar supremo se toma ínfimo. Es de notar que si f es cóncava, f^* no está definida como $-(-f)^*$, el siguiente teorema aclara esto.

Teorema 1.9. Sea $g : \mathbb{R} \rightarrow [-\infty, +\infty)$ una función cóncava y definamos $f = -g$.
Entonces $g^*(s) = -f^*(-s)$

Demostración

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

□

Teorema 1.10 (ver [6], **Proposición X.1.3.1**). Sean

$$f : \mathbb{R}^n \rightarrow (-\infty, +\infty], f_1, \dots, f_m : \mathbb{R} \rightarrow (-\infty, +\infty]$$

funciones convexas propias cerradas.

1. Si $\alpha \in \mathbb{R}$ y $h(x) = f(x) + \alpha$, entonces $h^*(s) = f^*(s) - \alpha$.
2. Si $\alpha > 0$ y $h(x) = \alpha f(x)$, entonces $h^*(s) = \alpha f^*\left(\frac{s}{\alpha}\right)$.
3. Si $\alpha \neq 0$ y $h(x) = f(\alpha x)$, entonces $h^*(s) = f^*\left(\frac{s}{\alpha}\right)$.
4. Si $y \in \mathbb{R}^n$ y $h(x) = f(x - y)$, entonces $h^*(s) = f^*(s) + sy$.
5. Si $x = (x_1, \dots, x_n)$ y $h(x) = \sum_{j=1}^m f_j(x_j)$, entonces $h^*(s) = \sum_{j=1}^m f_j^*(s_j)$ para $s = (s_1, \dots, s_m)$.

La siguiente definición, es una especie de extensión del concepto de gradiente, para el caso de funciones convexas no necesariamente diferenciables.

Definición 1.13. Sea $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ una función convexa. Un vector $s \in \mathbb{R}^n$ es un subgradiente de f en $x \in \mathbb{R}^n$ si

$$\forall y \in \mathbb{R}^n : f(y) \geq f(x) + s(y - x).$$

El conjunto formado por todos los vectores subgradientes de f en x , denotado por $\partial f(x)$, es llamado el subdiferencial de f en x .

Para una función cóncava $f : \mathbb{R}^n \rightarrow [-\infty, +\infty)$ las definiciones son análogas, sólo hay que invertir la desigualdad.

Si f es diferenciable en x , entonces $\partial f(x) = \{\nabla f(x)\}$

Teorema 1.11 (ver [6], Corolario X.1.4.4). Sea $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ una función convexa propia cerrada. Entonces para todo $s, x \in \mathbb{R}^n$ se cumple que

$$x \in \partial f^*(s) \Leftrightarrow s \in \partial f(x) \Leftrightarrow f(x) + f^*(s) - sx = 0.$$

Teorema 1.12 (Dualidad de Fenchel)(ver [10], Teorema 31.1). Supongamos que $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ es una función convexa propia y que $g : \mathbb{R}^n \rightarrow [-\infty, +\infty)$ es una función cóncava propia. Si algunas de las siguientes condiciones se cumple

(a) $\text{ri}(\text{dom}f) \cap \text{ri}(\text{dom}g) \neq \emptyset$,

(b) f y g son cerradas y $\text{ri}(\text{dom}f^*) \cap \text{ri}(\text{dom}g^*) \neq \emptyset$,

entonces

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

Además, si (a) se satisface, el supremo es alcanzado en algún s ; si (b) se satisface, el ínfimo es alcanzado en algún x ; si (a) y (b) son satisfechas simultáneamente, el ínfimo y el supremo son finitos.

Capítulo 2

Métodos de Multiplicadores y Punto Proximal

En este capítulo presentaremos algunos resultados ya estudiados sobre los métodos de multiplicadores y punto proximal, mayormente extraídas de [7]. Haremos algunas observaciones sobre la ventajas y desventajas de cada uno y sobre la relación de dualidad entre ellos.

En lo que resta del trabajo, usaremos las siguientes notaciones

- (i) $\mathcal{P}(\mathbb{R}^m)$ representa el conjunto de partes o conjunto potencia de \mathbb{R}^m .
- (ii) Dado $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, diremos que $x > 0$ si y sólo si $x_i > 0$ para todo $i \in \{1, \dots, n\}$.
- (iii) Dado $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, diremos que $x \geq 0$ si y sólo si $x_i \geq 0$ para todo $i \in \{1, \dots, n\}$.
- (iv) $\mathbb{R}_+^n = \{x \in \mathbb{R}^n : x > 0\}$.
- (v) $\mathbb{R}_{++}^n = \{x \in \mathbb{R}^n : x \geq 0\}$.
- (vi) Si $S \subset \mathbb{R}^n$, denotaremos por \overline{S} y ∂S a la clausura y la frontera de S respectivamente.
- (vii) Si $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ es una función diferenciable, denotaremos por $\nabla_1 f(x, y)$ y $\nabla_2 f(x, y)$ a los gradientes de f tomados respecto a la primera y segunda variable respectivamente.

2.1. Métodos de Punto Proximal

Nuestro interés principal, es aplicar este método para resolver un problema de minimización de una función convexa definida en \mathbb{R}_+^m , usando una cuasidistancia que definiremos en el capítulo 3.

2.1.1. El Método de Punto Proximal Clásico

Consideremos el problema

$$\inf\{f(x) : x \in \mathbb{R}^m\} \quad (2.1.1)$$

donde $f : \mathbb{R}^m \rightarrow \mathbb{R}$ es una función convexa propia cerrada.

El método de punto proximal clásico (ver [7]), aplicado a este problema, genera una sucesión $\{x^k\}$ mediante el siguiente algoritmo

$$x^0 \in \mathbb{R}^m \quad (2.1.2)$$

$$x^{k+1} = \operatorname{argumín}\{f(x) + \lambda_k \|x - x^k\|^2 : x \in \mathbb{R}^m\} \quad (2.1.3)$$

donde $\lambda_k \in \mathbb{R}$ y $0 < \lambda_k \leq \widehat{\lambda}$ para algún $\widehat{\lambda} > 0$.

Quisieramos que la sucesión generada por (2.1.2), (2.1.3) converja a una solución del problema (2.1.1), esto es posible bajo ciertas hipótesis específicas.

2.1.2. Método de Punto Proximal con Operadores Monótonos Maximales

Seguidamente daremos algunas definiciones y resultados ya probados en [7] en relación al método de punto proximal usando operadores maximales monótonos.

Definición 2.1. Un operador punto a conjunto $T : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$ es monótono si

$$(x - y)(u - v) \geq 0, \quad \forall x, y \in \mathbb{R}^m, \quad \forall u \in T(x), \quad \forall v \in T(y).$$

Ejemplo 2.1 (ver [7]). Si $f : \mathbb{R}^m \rightarrow \mathbb{R}$ es una función convexa, entonces $T = \partial f$ es un operador monótono.

En efecto, sean $x, y \in \mathbb{R}^m$, $u \in \partial f(x)$ y $v \in \partial f(y)$ cualesquiera, de la definición de subgradiente de una función convexa, se tiene

$$u(y - x) \leq f(y) - f(x)$$

y

$$v(x - y) \leq f(x) - f(y),$$

por lo tanto

$$(x - y)(u - v) = -u(y - x) - v(x - y) \geq 0.$$

□

Definición 2.2. Un operador $T : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$ es monótono maximal si

(i) T es monótono.

(ii) Para todo operador monótono \widehat{T} tal que $T(x) \subset \widehat{T}(x)$ para todo x , se cumple que $T = \widehat{T}$.

Ejemplo 2.2 (ver [7]). Sea $f : \mathbb{R}^m \rightarrow \mathbb{R}$ una función convexa. Se puede probar que ∂f es un operador monótono maximal. Este resultado es importante para el estudio que haremos más adelante. □

Como lo afirma Iusem en [7], el concepto de operador monótono maximal, dado en la definición anterior, viene a generalizar el concepto de transformación lineal semidefinida positiva para el caso no lineal. En el ejemplo anterior vemos que el subgradiente de una función monótona es un operador monótono maximal, lo cual es una ventaja para nuestro posterior estudio.

Definición 2.3. Un operador $T : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$ es llamado no expansivo si

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

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

Definición 2.4. Un operador $T : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$ es llamado firmemente no expansivo si

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

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

Definición 2.5. Una sucesión $\{y^k\}$ es Fejér convergente a un conjunto $U \subset \mathbb{R}^m$ con respecto a la norma Euclidiana si

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

Teorema 2.1 (ver [7], **Proposición 4.1**). Si T es firmemente no expansivo, entonces la sucesión definida por

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

$$x^{k+1} = T(x^k)$$

es Fejér convergente al conjunto de puntos fijos de T

Observación 2.1 (ver [7]). Dado un operador punto a conjunto $T : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$, si suponemos que $T(x) \neq \emptyset$ para algún $x \in \mathbb{R}^m$, se puede definir un nuevo operador $T^{-1} : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$ (una especie de operador inverso de T) como sigue

$$y \in T^{-1}(x) \Leftrightarrow x \in T(y).$$

Definición 2.6. Diremos que $x \in \mathbb{R}^m$ es un cero del operador monótono maximal $T : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$ si $0 \in T(x)$.

Volvamos a nuestro problema original dado por (2.1.1). Se tiene que

$$\begin{aligned} 0 \in \partial f(x) &\Leftrightarrow \forall y : 0 = 0(y - x) \leq f(y) - f(x) \\ &\Leftrightarrow \forall y : f(x) \leq f(y), \end{aligned}$$

es decir, x es un cero de ∂f si y sólo si x es una solución de (2.1.1), así que, el problema de encontrar ceros de un operador monótono maximal, generaliza el problema de minimización de funciones convexas.

Usando la condición de optimalidad en (2.1.3) obtenemos

$$0 \in \partial f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k)$$

que es equivalente a

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

por lo que en el contexto de operadores monótonos maximales sería natural escribir

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

o bien

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

es decir,

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

El siguiente teorema (ver [7], Teorema 4.2) garantiza que la sucesión $\{x^k\}$ definida por

$$x^0 \in \mathbb{R}^m \tag{2.1.4}$$

$$x^{k+1} \in (I + \frac{1}{\lambda_k}T)^{-1}(x^k) \tag{2.1.5}$$

con $0 < \lambda_k \leq \hat{\lambda}$, converge a un cero del operador $(I + \frac{1}{\lambda_k}T)^{-1}$. Al tomar $T = \partial f$, obtenemos la convergencia del algoritmo de punto proximal dado en (2.1.2), (2.1.3).

Teorema 2.2 (ver [7] Teorema 4.2). Si $T : \mathbb{R}^m \rightarrow \mathcal{P}(\mathbb{R}^m)$ es un operador monótono maximal y existe \bar{x} tal que $0 \in T(\bar{x})$, entonces la sucesión definida por (2.1.4), (2.1.5) converge a un vector x^* tal que $0 \in T(x^*)$.

2.1.3. Método de Punto Proximal con Cuasidistancias

Existen versiones del método de punto proximal usando las llamadas *cuasidistancias*. Básicamente se sustituye el núcleo cuadrático $\|x - x^k\|^2$ en el algoritmo original por $D(x, x^k)$ donde D es una cuasidistancia.

Definición 2.7. Sea X un subconjunto no vacío de \mathbb{R}^m . Una función $D : X \times X \rightarrow \mathbb{R}$ es llamada una cuasidistancia en X si

$$(i) \quad \forall x, y \in X : D(x, y) \geq 0;$$

$$(ii) \quad D(x, y) = 0 \Leftrightarrow x = y.$$

Entre las cuasidistancias que se han usado para generalizar el método de punto proximal, están las llamadas *distancias de Bregman* y las φ -*divergencias*, ambas son casos particulares de las llamadas *medidas divergentes*.

Definición 2.8. Dado $S \subset \mathbb{R}^m$, una función $d : S \times S \rightarrow \mathbb{R}$ es llamada una medida divergente en S si

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

$$(ii) \quad \text{Si } \{x^k\} \subset S \text{ y } x \in S, \text{ entonces } \lim_{k \rightarrow +\infty} d(x, x^k) = 0 \Leftrightarrow \lim_{k \rightarrow +\infty} x^k = x;$$

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

(iv) Los conjuntos de nivel $\Gamma_2(x, \nu) = \{y \in S : d(x, y) \leq \nu\}$ son acotados para todo $x \in S$ y todo $\nu > 0$.

La condición (ii) implica que

$$d(x, y) = 0 \Leftrightarrow x = y,$$

es decir, toda medida divergente es una cuasidistancia. Un caso particular de las medidas divergentes son las llamadas φ -divergencias.

Definición 2.9. Sea $\varphi : \mathbb{R}_{++} \rightarrow \mathbb{R}$ una función continua tal que

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

(ii) $\varphi(1) = 0$;

(iii) $\lim_{t \rightarrow +\infty} \varphi(t) = +\infty$.

Definamos la función $d_\varphi : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ como

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

La función d_φ es llamada φ -divergencia.

Un estudio detallado del algoritmo de punto proximal usando este tipo de funciones se puede encontrar en [7], [8] y [9].

Otra caso particular de las medidas divergentes son las llamadas distancias de Bregman.

Definición 2.10. Sea $S \subset \mathbb{R}^n$ un conjunto abierto convexo. Consideremos una función convexa $h : \bar{S} \rightarrow \mathbb{R}$ diferenciable en S . Definamos la función $D_h : \bar{S} \times S \rightarrow \mathbb{R}$ por

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

h es llamada una función de Bregman con zona S (y D_h la distancia de Bregman inducida por h) si se cumplen las siguientes condiciones

(B1) h es continuamente diferenciable en S .

(B2) h es estrictamente convexa y continua en \bar{S} .

(B3) Para todo $\delta \in \mathbb{R}$ los conjuntos de nivel $\Gamma_1(y, \delta) = \{x \in \bar{S} : D_h(x, y) \leq \delta\}$
 $\Gamma_2(x, \delta) = \{y \in S : D_h(x, y) \leq \delta\}$ son acotados para todo $y \in S$ y todo $x \in \bar{S}$
respectivamente.

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

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

Se prueba que toda distancia de Bregman es una cuasidistancia. (B4) y (B5) son consecuencia de (B1), (B2) y (B3) si $\{x^k\} \subset S$ y $y^* \in S$, así que (B4) y (B5) sólo necesitan ser verificadas si $\{x^k\} \subset \partial S$ o $y^* \in \partial S$.

Definición 2.11. Una función de Bregman h con zona S es llamada coerciva en la frontera cuando satisface que

(B6) Si $\{y^k\} \subset S$ es tal que $\lim_{k \rightarrow +\infty} y^k = y \in \partial S$, entonces

$$\lim_{k \rightarrow +\infty} \nabla h(y^k)(x - y^k) = -\infty.$$

Definición 2.12. Una función de Bregman h con zona S es llamada coerciva en la zona cuando satisface que

(B7) Para todo $y \in \mathbb{R}^n$ existe $x \in S$ tal que $\nabla h(x) = y$.

Un estudio del método de punto proximal usando distancias de Bregman es hecho en [7].

Sea $S \subset \mathbb{R}^m$ un conjunto abierto y $f : \mathbb{R}^m \rightarrow \mathbb{R}$ una función convexa en \bar{S} . Consideremos el problema

$$\inf\{f(x) : x \in \bar{S}\} \tag{2.1.6}$$

El algoritmo de punto proximal aplicado a este problema, usando una cuasidistancia $D : S \times S \rightarrow \mathbb{R}$, genera una sucesión $\{x^k\}$ dada por

$$x^0 \in S \tag{2.1.7}$$

$$x^{k+1} = \operatorname{argumín}\{f(x) + \lambda_k D(x, x^k) : x \in S\}. \tag{2.1.8}$$

donde $\lambda_k \in \mathbb{R}$ y $0 < \lambda_k \leq \hat{\lambda}$ para algún $\hat{\lambda} > 0$.

Bajo ciertas hipótesis se puede probar que la sucesión dada por (2.1.7) y (2.1.8) converge a una solución óptima de (2.1.6).

2.2. Métodos de Lagrangeano Aumentado

En la sección anterior estudiamos el algoritmo de punto proximal. Cabe destacar que este tipo de métodos no es eficaz computacionalmente, el hecho es que cada iteración del método requiere una minimización de una función en \mathbb{R}^m . Sin embargo, resultan muy útiles como una herramienta conceptual en el análisis de convergencia de algoritmos en problemas de optimización.

En esta sección estudiaremos los métodos de lagrangeano aumentado, que contrariamente a los métodos de punto proximal, resultan una eficaz herramienta computacional, más no teórica, en el estudio de convergencia de los algoritmos antes mencionados. Estos métodos sustituyen un problema de minimización con restricciones, por una secuencia de problemas irrestrictos. Además, están estrechamente relacionados, como veremos luego, con los métodos de punto proximal vía dualidad de Fenchel.

2.2.1. El Lagrangeano Clásico

Consideremos el problema de programación convexa

$$(P) \quad \begin{array}{ll} \text{minimizar} & f_0(x) \\ \text{sujeto a} & f_i(x) \leq 0, \quad i \in \{1, \dots, m\}. \end{array}$$

donde $f_0, f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$ son funciones convexas propias cerradas. El *lagrangeano clásico* $l : \mathbb{R}^n \times \mathbb{R}_{++}^m \rightarrow \mathbb{R}$ para éste problema está dado por

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

El método de lagrangeano clásico genera dos sucesiones $\{x^k\}$ y $\{\mu^k\}$ mediante el siguiente algoritmo, comenzando con $\mu^0 \in \mathbb{R}_{++}^m$ definimos

$$x^{k+1} = \operatorname{argmín}\{l(x, \mu^k) : x \in \mathbb{R}^n\} \quad (2.2.1)$$

la actualización para μ^k no es especificada, sin embargo, debe hacerse de manera apropiada.

Las hipótesis que comunmente se dan para este tipo de problemas son las siguientes

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

(A2) *Existe $\hat{x} \in \operatorname{dom} f_0$ tal que $f_i(\hat{x}) < 0$ para $i = 1, \dots, m$ (Condición de Calificación de Slater).*

El *problema dual* del problema (P), el cual sabemos que es un problema cóncavo, viene dado por:

$$(D) \quad \sup\{d(\mu) : \mu \in \mathbb{R}_+^m\}$$

donde

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

es la función dual objetivo del problema (P).

Observación 2.2. (A2) implica que el conjunto de soluciones óptimas del problema (D) es compacto y no vacío (ver [10] secciones 28 y 29) y que $\widehat{f} = \widehat{d}$ donde

$$\widehat{f} = \inf\{f_0(x) : x \in \mathbb{R}^n, f_i(x) \leq 0, i = 1, \dots, m\}$$

y

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

Además, para cada $\beta < \widehat{d}$, el conjunto de nivel

$$\{\mu \in \mathbb{R}_+^m : d(\mu) \geq \beta\}$$

es compacto. □

2.2.2. Lagrangeano Aumentado

Consideremos el problema (P) dado en la sección anterior. El *lagrangeano aumentado clásico* $L_\alpha : \mathbb{R}^n \times \mathbb{R}_{++}^m \rightarrow \mathbb{R}$ para este problema está dado por

$$L_\alpha(x, \mu) = f_0(x) + \frac{1}{4\alpha} \sum_{i=1}^m \{[(\mu_i + 2\alpha f_i(x))_+]^2 - \mu_i^2\}$$

donde $\alpha > 0$ y $a_+ = \max\{0, a\}$

Esta función es diferenciable y además

$$\nabla_1 L_\alpha(x, \mu) = \nabla f_0(x) + \sum_{i=1}^m (\mu_i + 2\alpha f_i(x))_+ \nabla f_i(x)$$

Al aplicar el método del lagrangeano aumentado al problema (P), se obtienen dos sucesiones $\{x^k\}$, $\{\mu^k\}$ de la siguiente manera. Escojamos $\mu^0 \in \mathbb{R}_{++}^n$ y definamos

$$x^{k+1} = \operatorname{argumín}\{L_\alpha(x, \mu^k) : x \in \mathbb{R}^n\} \quad (2.2.2)$$

$$\mu_i^{k+1} = (\mu_i^k + 2\alpha f_i(x^{k+1}))_+, \text{ para } i = 1, \dots, m. \quad (2.2.3)$$

Al aplicar la condición de optimalidad a x^{k+1} se obtiene

$$\nabla f_0(x^{k+1}) + \sum_{i=1}^m (\mu_i^k + 2\alpha f_i(x^{k+1}))_+ \nabla f_i(x^{k+1}) = 0$$

y al usar (2.2.3) nos queda

$$\nabla f_0(x^{k+1}) + \sum_{i=1}^m \mu_i^{k+1} \nabla f_i(x^{k+1}) = 0$$

es decir, x^{k+1}, μ^{k+1} satisfacen la condición de optimalidad para el lagrangeano clásico, de aquí el por qué de la actualización en (2.2.3), ya que esta es una condición necesaria para que x^{k+1} sea solución (no necesariamente óptima) de (P) .

Otros tipos de lagrangeano aumentado surgen al considerar las llamadas funciones de penalidad. Por ejemplo, dado $r \in (0, 1]$ y una función creciente, diferenciable y estrictamente convexa $\theta : (-\infty, a) \rightarrow \mathbb{R}$ donde $a \in (0, +\infty]$ tal que

- i. $\theta(0) = 0$;
- ii. $\theta'(0) = 1$;
- iii. $\lim_{t \rightarrow a^-} \theta'(t) = +\infty$;
- iv. $\lim_{t \rightarrow -\infty} \theta'(t) = 0$.

Las siguientes funciones pueden definirse como el *lagrangeano aumentado* del problema (P) .

1. $L_r : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ dada por $L_r(x, \mu) = f_0(x) + r \sum_{i=1}^m \mu_i \theta\left(\frac{f_i(x)}{r}\right)$ (ver [9]).
2. $L_r : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ dada por $L_r(x, \mu) = f_0(x) + r \sum_{i=1}^m (\mu_i)^2 \theta\left(\frac{f_i(x)}{r\mu_i}\right)$ (ver [1]).

Los respectivos algoritmos que se obtienen al aplicar el método de lagrangeano aumentado se obtienen al sustituir $L_\alpha(x, \mu^k)$ en (2.2.2) por $L_r(x, \mu^k)$ y

$$\mu_i^{k+1} = (\mu_i^k + 2\alpha f_i(x^{k+1}))_+$$

en (2.2.3) por

$$\mu_i^{k+1} = \mu_i^k \theta' \left(\frac{f_i(x^{k+1})}{r} \right) \text{ para (1);}$$

$$\mu_i^{k+1} = \mu_i^k \theta' \left(\frac{f_i(x^{k+1})}{r\mu_i^k} \right) \text{ para (2).}$$

Estas últimas actualizaciones permiten probar, al igual que para el caso del lagrangeano aumentado clásico, que μ^{k+1}, x^{k+1} satisfacen la condición de optimalidad para el lagrangeano clásico, de aquí el porqué se toman de esta manera.

Bajo hipótesis específicas, que pueden diferir en cada caso, se puede probar que los puntos límites de la sucesión $\{x^k\}$, o bien la sucesión de promedios de esta última, converge a una solución óptima del problema (P) y que cualquier punto límite de la sucesión $\{\mu^k\}$, o la sucesión completa, converge a una solución óptima del problema (D) .

Capítulo 3

Método Propuesto

En [9] Polyak y Teboulle presentan una metodología en métodos de multiplicadores que denominan “Nonlinear Rescaling” (NR), la cual consiste, básicamente, en transformar el problema (P) en uno equivalente cuya función lagrangeano $L_r(x, \mu) : \mathbb{R}^n \times \mathbb{R}_{++}^m$ viene dada por

$$L_r(x, \mu) = f_0(x) + r \sum_{j=1}^m \mu_j \theta \left(\frac{f_j(x)}{r} \right) \quad (3.0.4)$$

donde r es un número real positivo, y θ es una función de penalidad. La cuasidistancia asociada a $L_r(x, \mu)$ vía conjugacidad, es una φ -divergencia. Polyak y Teboulle presentan, en primer lugar, un método de multiplicadores para resolver (P) , generando un método de punto proximal que usan para resolver (D) . La sucesión primal generada por el método converge de manera ergódica a una solución óptima de (P) , esto es, la sucesión de promedios de la sucesión primal converge a una solución óptima de (P) , en cambio, la sucesión de puntos duales converge de manera global a una solución óptima de (D) .

Ben-Tal y Zibulevsky en [3], al igual que Polyak y Teboulle en [9], presentan en primer lugar un método de multiplicadores que llaman “The Penalty/Barrier Multiplier Method” (PBM) que usan para resolver (P) . Al considerar (D) , se encuentran con un método de punto proximal cuyo núcleo es una cuasidistancia que involucra una función sublineal. Al hacer el estudio de convergencia, obtienen que cualquier punto límite de la sucesión primal converge a una solución óptima de (P) y cualquier punto límite de la sucesión dual converge a una solución óptima de (D) . La función

lagrangeano aumentado $L(x, \mu, p) : \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ usada por ellos viene dada por:

$$L(x, \mu, p) = f_0(x) + \sum_{j=1}^m \mu_j p_j \theta \left(\frac{f_j(x)}{p_j} \right) \quad (3.0.5)$$

con θ una función de penalidad. Es de hacer notar que al tomar $p = r\mu$ en (3.0.5), con r un número real positivo, se genera, en el núcleo del método de punto proximal asociado, una cuasidistancia homogénea de segundo orden.

Auslender, Teboulle y Ben-Tiba en [1], a diferencia de Polyak y Teboulle en [9] y Ben-Tal y Zibulevsky en [3], comienzan generando un algoritmo que llaman “Basic Iterations Scheme” (BIS), que es una versión aproximada del método de punto proximal, el cual usa una cuasidistancia homogénea de segundo orden y es usado para resolver (D) . Se estudian dos clases de algoritmos basados en BIS, estos son “Interior Proximal Method” (IPM) y “Regularized Interior Proximal Method” (RIPM). Al considerar (P) vía dualidad, utilizan un método de multiplicadores inexactos. En el análisis de convergencia se obtienen dos resultados para los dos algoritmos antes mencionados. La correspondiente función lagrangeano aumentado $L_r(x, \mu) : \mathbb{R}^n \times \mathbb{R}_{++}^m$ es dada por:

$$L_r(x, \mu) = f_0(x) + r \sum_{j=1}^m (\mu_j)^2 \theta \left(\frac{f_j(x)}{r\mu_j} \right) \quad (3.0.6)$$

donde r es un número real positivo y θ una función de penalidad.

La función dada en (3.0.6), es un caso particular de la dada en (3.0.5), sin embargo, los resultados dados por Auslender, Teboulle y Ben-Tiba en [1] se obtienen por una vía distinta a los resultados obtenidos por Ben-Tal y Zibulevsky en [3].

Al observar las funciones lagrangeano aumentado dadas por (3.0.4), (3.0.5) y (3.0.6), surge la interrogante en el sentido de analizar si es o no posible considerar potencias cúbicas de los multiplicadores que aparecen en las funciones objetivos y qué consideraciones y resultados adicionales se podrían obtener.

En esta dirección, hemos formulado nuestro principal objetivo, el cual consiste en

obtener un método de multiplicadores alternativo, en este caso, utilizando potencias cúbicas para los multiplicadores, que nos permita resolver (P) .

A continuación presentaremos este nuevo método de multiplicadores que genera, vía dualidad de Fenchel, un método de punto proximal, que usaremos para resolver el problema (D) , y que tiene como núcleo una cuasidistancia homogénea de tercer orden. Seguidamente estudiaremos un algoritmo generado por este método, este estudio incluirá, entre otras cosas, un análisis de convergencia del mismo. Además, aunque nuestro objetivo principal es el estudio teórico del método y no su estudio numérico, haremos una implementación en MATLAB del método al final del capítulo e incluiremos algunos ejemplos, escogidos de manera adecuada, que resolveremos usando esta implementación.

3.1. Desarrollo del Método

Consideremos el problema de programación convexa

$$(P) \quad \begin{array}{ll} \text{minimizar} & f_0(x) \\ \text{sujeto a} & f_i(x) \leq 0, \quad i \in \{1, \dots, m\}. \end{array}$$

donde f_0, f_1, \dots, f_m son funciones convexas de clase C^1 en \mathbb{R}^n . Los resultados que se van a derivar en este trabajo continúan siendo válidos en el caso de que las funciones f_0, f_1, \dots, f_m sean convexas propias cerradas (no necesariamente diferenciables).

Supongamos además que se cumplen las condiciones $(A1)$ y $(A2)$ dadas en el capítulo anterior.

Consideremos además el problema dual del problema (P)

$$(D) \quad \begin{array}{ll} \text{maximizar} & d(\mu) \\ \text{sujeto a} & \mu \geq 0 \end{array}$$

donde d es la función dual objetivo del problema (P) ya definida en el capítulo anterior.

Definamos una clase de funciones que nos ayudará a establecer el método. Sea θ una función de clase C^2 en el intervalo $(-\infty, a)$ con $0 < a \leq +\infty$, la cual satisface las siguientes propiedades

($\theta 1$) θ es estrictamente creciente, esto es, $\theta'(t) > 0$ para todo $t \in (-\infty, a)$,

($\theta 2$) θ es estrictamente convexa, es decir, $\theta''(t) > 0$ para $t \in (-\infty, a)$,

($\theta 3$) $\theta(0) = 0$, $\theta'(0) = 1$,

($\theta 4$) $\lim_{t \rightarrow a^-} \theta'(t) = +\infty$,

($\theta 5$) $\lim_{t \rightarrow -\infty} \theta'(t) = 0$,

($\theta 6$) Existe $M > 0$ tal que $\theta''(t) \geq M$ para todo $t \in [0, a]$.

Llamemos Θ a la clase de funciones que cumplen con estas propiedades.

Observación 3.1. Dada $\theta \in \Theta$, ($\theta 2$) garantiza que θ' es creciente, en consecuencia

$$t < 0 \Rightarrow \theta'(t) \leq \theta'(0) = 1 \Rightarrow t\theta'(t) \geq t$$

$$0 < t < a \Rightarrow \theta'(t) \geq \theta'(0) = 1 \Rightarrow t\theta'(t) \geq t.$$

Es decir, $t\theta'(t) \geq t$ para todo $t \in (-\infty, a)$. □

Para $\theta \in \Theta$ y $r > 0$ definamos la función

$$L_r(x, \mu) = f_0(x) + r \sum_{i=1}^m (\mu_i)^3 \theta \left(\frac{f_i(x)}{r(\mu_i)^2} \right).$$

Llamaremos a esta función el *lagrangeano aumentado* del problema (P). No se debe confundir esta función con el lagrangeano clásico $l(x, \mu)$.

Considerando esta función, formaremos el siguiente algoritmo para resolver el problema (P).

Algoritmo. Dados $\theta \in \Theta$, $\mu^0 \in \mathbb{R}_{++}^m$ y $r > 0$, generamos iterativamente las sucesiones $\{x^k\}$ y $\{\mu^k\}$ como sigue

$$x^{k+1} \in \operatorname{argmín}\{L_r(x, \mu^k) : x \in \mathbb{R}^n\}, \quad (3.1.1)$$

$$\mu_i^{k+1} = \mu_i^k \theta' \left(\frac{f_i(x^{k+1})}{r(\mu_i^k)^2} \right), \quad i = 1, \dots, m. \quad (3.1.2)$$

Se puede variar el valor r en cada iteración bajo la condición de que estos valores estén en un intervalo cerrado $[\tilde{r}, \hat{r}]$ con $\tilde{r} > 0$.

Usando la Proposición 1 dada en [3], aplicada al problema $\min_x L_r(x, \mu)$, se garantiza que el mínimo en el paso (3.1.1) es alcanzado y en consecuencia el algoritmo está bien definido.

De la condición de optimalidad para x^{k+1} tenemos que:

$$\nabla f_0(x^{k+1}) + \sum_{i=1}^m \mu_i^k \theta' \left(\frac{f_i(x^{k+1})}{r(\mu_i^k)^2} \right) \nabla f_i(x^{k+1}) = 0. \quad (3.1.3)$$

Sustituyendo (3.1.2) en (3.1.3) obtenemos

$$\nabla f_0(x^{k+1}) + \sum_{i=1}^m \mu_i^{k+1} \nabla f_i(x^{k+1}) = 0. \quad (3.1.4)$$

Es decir, x^{k+1} , μ^{k+1} satisfacen la condición de optimalidad para la minimización del lagrangeano clásico, $\nabla_1 l(x^{k+1}, \mu^{k+1}) = 0$. Más aún

$$\hat{f} = \hat{d} \geq d(\mu^{k+1}) = \inf_x l(x, \mu^{k+1}) = l(x^{k+1}, \mu^{k+1}). \quad (3.1.5)$$

Además, dado que $\mu^0 > 0$, por $(\theta 1)$ y $(\theta 5)$ tenemos que la sucesión generada por (3.1.2) satisface que $\mu^k > 0$ para todo k , y por lo tanto son puntos factibles del problema (D) . Sin embargo, la sucesión generada por (3.1.1) no necesariamente genera puntos factibles del problema (P) . Pero al final el capítulo veremos que esta última sucesión satisface que dado $\varepsilon > 0$

$$f_i(x^k) < \varepsilon$$

para todo $i = 1, \dots, m$ y para k suficientemente grande, es decir, si bien no se puede garantizar la factibilidad de los puntos de la sucesión, al menos se puede garantizar que son “asintóticamente factibles”. Además, veremos que los puntos límites de las sucesiones dadas por (3.1.1) y (3.1.2), son soluciones de (P) y (D) respectivamente.

De (3.1.5) se tiene que

$$\begin{aligned} d(\mu) - d(\mu^{k+1}) &= \inf_x l(x, \mu) - \inf_x l(x, \mu^{k+1}) \\ &\leq l(x^{k+1}, \mu) - l(x^{k+1}, \mu^{k+1}) \\ &= \sum_{i=1}^m (\mu_i - \mu_i^{k+1}) f_i(x^{k+1}) \\ &= (f_1(x^{k+1}), \dots, f_m(x^{k+1}))(\mu - \mu^{k+1}). \end{aligned}$$

Por lo tanto

$$(f_1(x^{k+1}), \dots, f_m(x^{k+1}))^T \in \partial d(\mu^{k+1}). \quad (3.1.6)$$

Por (θ2) se tiene que la función $(\theta')^{-1}$ existe. Ahora bien, para todo $i = 1, \dots, m$

$$\begin{aligned} \mu_i^{k+1} = \mu_i^k \theta' \left(\frac{f_i(x^{k+1})}{r(\mu_i^k)^2} \right) &\Rightarrow \frac{\mu_i^{k+1}}{\mu_i^k} = \theta' \left(\frac{f_i(x^{k+1})}{r(\mu_i^k)^2} \right) \\ &\Rightarrow f_i(x^{k+1}) = r(\mu_i^k)^2 (\theta')^{-1} \left(\frac{\mu_i^{k+1}}{\mu_i^k} \right) \\ &\Rightarrow f_i(x^{k+1}) = r(\mu_i^k)^2 (\theta^*)' \left(\frac{\mu_i^{k+1}}{\mu_i^k} \right) \end{aligned}$$

en la última igualdad usamos la relación $(\theta')^{-1} = (\theta^*)'$, donde $\theta^*(s) = \sup_t \{st - \theta(t)\}$ es la conjugada de θ , obteniendo

$$0 \in \partial d(\mu^{k+1}) - r \left((\mu_1^k)^2 (\theta^*)' \left(\frac{\mu_1^{k+1}}{\mu_1^k} \right), \dots, (\mu_m^k)^2 (\theta^*)' \left(\frac{\mu_m^{k+1}}{\mu_m^k} \right) \right)^T. \quad (3.1.7)$$

Proposición 3.1. *Sea $\theta \in \Theta$. Entonces la función $\varphi = \theta^*$, la función conjugada de θ , cumple con las siguientes propiedades:*

(φ1) φ es una función diferenciable, estrictamente convexa en $(0, +\infty)$,

$$(\varphi 2) \quad \varphi(1) = 0,$$

$$(\varphi 3) \quad \varphi'(1) = 0,$$

$$(\varphi 4) \quad \lim_{s \rightarrow 0^+} \varphi'(s) = -\infty,$$

$$(\varphi 5) \quad \lim_{s \rightarrow +\infty} \varphi'(s) = a,$$

$$(\varphi 6) \quad \varphi''(s) \leq \frac{1}{M} \text{ para } s \geq 1.$$

Demostración

Sólo probaremos las propiedades $(\varphi 5)$ y $(\varphi 6)$, para una prueba del resto de las propiedades ver [9] (Proposición 3.1).

De $(\theta 4)$

$$\lim_{s \rightarrow +\infty} \varphi'(s) = \lim_{s \rightarrow +\infty} (\theta')^{-1}(s) = a,$$

lo cual prueba $(\varphi 5)$.

Por otro lado

$$\varphi''(s) = [(\theta')^{-1}]'(s) = \frac{1}{\theta''[(\theta')^{-1}(s)]} = \frac{1}{\theta''[\varphi'(s)]}. \quad (3.1.8)$$

De $(\varphi 1)$, $(\varphi 2)$ y $(\varphi 3)$, $\varphi'(s) \geq 0$ para $s \geq 1$, además, por $(\varphi 5)$ y dado que φ' es continua, $\varphi'(s) \leq a$, es decir,

$$s \geq 1 \Rightarrow 0 \leq \varphi'(s) \leq a. \quad (3.1.9)$$

Finalmente, de (3.1.8), (3.1.9) y $(\theta 6)$, se tiene que $\varphi''(s) \leq \frac{1}{M}$ para $s \geq 1$. \square

Observación 3.2. Obsérvese que

$$\varphi(0) = \theta^*(0) = \sup_t \{-\theta(t)\} = -\inf_t \{\theta(t)\}$$

pero de $(\theta 1)$ y $(\theta 3)$ se tiene que $\theta(t) < 0$ para todo $t < 0$, así que

$$\varphi(0) > 0.$$

\square

Observación 3.3. (φ_1) , (φ_2) y (φ_3) implican que φ es decreciente en $(0, 1)$ y creciente en $(1, +\infty)$. \square

La inclusión (3.1.7) puede ser escrita como

$$0 \in \partial d(\mu^{k+1}) - r \left((\mu_1^k)^2 \varphi' \left(\frac{\mu_1^{k+1}}{\mu_1^k} \right), \dots, (\mu_m^k)^2 \varphi' \left(\frac{\mu_m^{k+1}}{\mu_m^k} \right) \right)^T. \quad (3.1.10)$$

Llamaremos Φ a la clase de funciones que satisfacen las propiedades en la proposición 3.1.

A continuación presentamos algunos ejemplos de funciones $\theta \in \Theta$ junto con sus conjugadas $\varphi \in \Phi$.

1. $\theta(t) = e^t - 1$, con $t \in \mathbb{R}$; $\varphi(s) = s \ln s - s + 1$.
2. $\theta(t) = \frac{t}{1-t}$, con $t < 1$; $\varphi(s) = (\sqrt{s} - 1)^2$.
3. $\theta(t) = -\ln(1-t)$, con $t < 1$; $\varphi(s) = s - 1 - \ln s$.
4. $\theta(t) = \frac{1}{2} \left[\frac{1}{(1-t)^2} - 1 \right]$, con $t < 1$; $\varphi(s) = \frac{(\sqrt[3]{s} - 1)^2 (2\sqrt[3]{s} + 1)}{2}$.
5. $\theta(t) = \frac{2}{\sqrt{1-t}} - 2$, con $t < 1$; $\varphi(s) = (\sqrt[3]{s} - 1)^2 (\sqrt[3]{s} + 2)$.

En la siguiente proposición, definiremos una cuasidistancia D_φ la cual es homogénea de tercer orden y va a ser de vital importancia en el estudio del problema dual (D), con esta cuasidistancia definiremos un método de punto proximal para resolver este problema.

Proposición 3.2 (ver [3] Proposición 2). Dada $\varphi \in \Phi$, definamos la función $D_\varphi : \mathbb{R}_+^m \times \mathbb{R}_{++}^m \rightarrow \mathbb{R}$ como

$$D_\varphi(x, y) = \sum_{i=1}^m (y_i)^3 \varphi \left(\frac{x_i}{y_i} \right).$$

Entonces

1. La función D_φ es homogénea de tercer orden, esto es

$$D_\varphi(\alpha x, \alpha y) = \alpha^3 D_\varphi(x, y) \text{ para todo } (x, y) \in \mathbb{R}_+^m \times \mathbb{R}_{++}^m \text{ y todo } \alpha > 0.$$

2. $D_\varphi(x, y) \geq 0$ para todo $(x, y) \in \mathbb{R}_+^m \times \mathbb{R}_{++}^m$.

3. $D_\varphi(x, y) = 0 \Leftrightarrow x = y$.

4. $D_\varphi(\cdot, y)$ es estrictamente convexa en \mathbb{R}_+^m para todo $y \in \mathbb{R}_{++}^m$.

5. $D_\varphi(\cdot, y)$ tiene conjuntos de nivel acotados para todo $y \in \mathbb{R}_{++}^m$.

Demostración

Sean $(x, y) \in \mathbb{R}_+^m \times \mathbb{R}_{++}^m$ y todo $\alpha > 0$. Entonces

$$D_\varphi(\alpha x, \alpha y) = \sum_{i=1}^m (\alpha y_i)^3 \varphi\left(\frac{\alpha x_i}{\alpha y_i}\right) = \alpha^3 \sum_{i=1}^m (y_i)^3 \varphi\left(\frac{x_i}{y_i}\right) = \alpha^3 D_\varphi(x, y).$$

Además $(\varphi 1)$, $(\varphi 2)$, $(\varphi 3)$ y la observación 3.2 garantizan que $\varphi(s) \geq 0$ para todo $s \geq 0$ y que $\varphi(s) = 0$ si y sólo si $s = 1$. Por lo tanto $D_\varphi(x, y) \geq 0$ para todo $(x, y) \in \mathbb{R}_+^m \times \mathbb{R}_{++}^m$ y además

$$\begin{aligned} D_\varphi(x, y) = 0 &\Leftrightarrow \sum_{i=1}^m (y_i)^3 \varphi\left(\frac{x_i}{y_i}\right) = 0 \\ &\Leftrightarrow (y_i)^3 \varphi\left(\frac{x_i}{y_i}\right) = 0, \forall i = 1, \dots, m \\ &\Leftrightarrow \varphi\left(\frac{x_i}{y_i}\right) = 0, \forall i = 1, \dots, m \\ &\Leftrightarrow \frac{x_i}{y_i} = 1, \forall i = 1, \dots, m \\ &\Leftrightarrow x = y. \end{aligned}$$

De $(\varphi 1)$ se tiene que para cada $\beta > 0$ fijo la función $h_\beta(\alpha) = \beta^3 \varphi\left(\frac{\alpha}{\beta}\right)$ es estrictamente convexa en $[0, +\infty)$. Así que para cada $y \in \mathbb{R}_{++}^m$ la función $D_\varphi(\cdot, y)$ es estrictamente convexa en \mathbb{R}_+^m .

Sean $N > 0$ y $y \in \mathbb{R}_{++}^m$ fijos. Entonces

$$\begin{aligned} D_\varphi(x, y) \leq N &\Rightarrow (y_i)^3 \varphi\left(\frac{x_i}{y_i}\right) \leq N, \forall i \in \{1, \dots, m\} \\ &\Rightarrow \varphi\left(\frac{x_i}{y_i}\right) \leq \frac{N}{(y_i)^3}, \forall i \in \{1, \dots, m\}. \end{aligned}$$

Pero los conjuntos

$$A_{i,N} = \left\{ t \geq 0 : \varphi\left(\frac{t}{y_i}\right) \leq \frac{N}{(y_i)^3} \right\}$$

son acotados, pues la función φ es continua. Por lo tanto $D_\varphi(\cdot, y)$ tiene conjuntos de nivel acotados para todo $y \in \mathbb{R}_{++}^m$. \square

3.2. Análisis de Convergencia

En esta sección se darán a conocer los resultados más importantes de nuestro trabajo, el estudio de la convergencia de las sucesiones generadas en (3.1.1) y (3.1.2). Comenzaremos por estudiar el problema (D) a través de un método de punto proximal.

La inclusión (3.1.10) puede ser reescrita en términos de la función D_φ como sigue

$$0 \in \partial d(\mu^{k+1}) - r \nabla_1 D_\varphi(\mu^{k+1}, \mu^k), \quad (3.2.1)$$

y esta última inclusión no es sino la condición de optimalidad para el problema cóncavo

$$\mu^{k+1} = \operatorname{argumáx}\{d(\mu) - r D_\varphi(\mu, \mu^k) : \mu \in \mathbb{R}_+^m\}. \quad (3.2.2)$$

Proposición 3.3 (ver [9] Proposición 4.1). *La sucesión $\{d(\mu^k)\}$ es no decreciente y acotada, y por lo tanto convergente.*

Demostración

Por (3.2.2), dado que $D_\varphi(x, y) \geq 0$ y $r > 0$, tenemos que

$$d(\mu^{k+1}) \geq d(\mu^{k+1}) - r D_\varphi(\mu^{k+1}, \mu^k) \geq d(\mu) - r D_\varphi(\mu, \mu^k), \forall \mu \geq 0$$

en particular, para $\mu = \mu^k$, pero $D_\varphi(\mu^k, \mu^k) = 0$, por lo tanto $d(\mu^{k+1}) \geq d(\mu^k)$, es decir, $\{d(\mu^k)\}$ es no decreciente. Además, por el Teorema de la Dualidad Debil (ver [2], Teorema 6.2.1), $d(\mu^k) \leq \widehat{f}$, en consecuencia $\{d(\mu^k)\}$ es convergente. \square

Proposición 3.4 (ver [9] **Proposición 4.2**). *La sucesión $\{\mu^k\}$ es acotada.*

Demostración

Por la observación 2.2 se tiene que el conjunto $\Lambda = \{\mu \in \mathbb{R}_+^m : d(\mu) \geq d(\mu^0)\}$ es compacto, además, por la proposición 3.3, $\mu^k \in \Lambda$ para todo k , por lo tanto $\{\mu^k\}$ es acotada. \square

De la proposición 3.4 se tiene que existe $\widehat{\mu} > 0$ tal que

$$0 < \mu^k \leq \widehat{\mu}, \forall k. \quad (3.2.3)$$

Proposición 3.5 (inspirada en [3] **Lema 2**). $\varphi(s_0) \geq \frac{M}{2}[\varphi'(s_0)]^2$ para todo $s_0 \geq 1$.

Demostración

Sea $s_0 \geq 1$. Definamos la función cuadrática

$$q(s) = (s - s_0)\varphi'(s_0) + \frac{(s - s_0)^2}{2M},$$

entonces

$$q'(s) = \varphi'(s_0) + \frac{s - s_0}{M} = 0 \Leftrightarrow s = s_0 - M\varphi'(s_0)$$

por lo tanto el minimizador de q es $s^* = s_0 - M\varphi'(s_0)$.

De la observación 3.3 y dado que $s_0 \geq 1$, se tiene que $\varphi'(s_0) \geq 0$ y así $s^* \leq s_0$.

Por otro lado, de $(\varphi 2)$ y en virtud del Teorema del Valor Medio

$$\varphi'(s_0) = \varphi'(s_0) - \varphi'(1) = (s_0 - 1)\varphi''(\bar{s})$$

para algún $\bar{s} \in [1, s_0]$.

De ($\varphi 6$)

$$\varphi''(\bar{s}) \leq \frac{1}{M}$$

Por lo tanto

$$\varphi'(s_0) \leq \frac{s_0 - 1}{M},$$

es decir,

$$1 \leq s_0 - M\varphi'(s_0) = s^*,$$

en consecuencia $1 \leq s^* \leq s_0$.

Usando nuevamente el Teorema del Valor Medio, se tiene que para todo $s \in [1, s_0]$

$$\varphi'(s_0) - \varphi'(s) = (s_0 - s)\varphi''(\hat{s})$$

para algún $\hat{s} \in [s, s_0]$.

De nuevo por ($\varphi 6$) y dado que $\hat{s} \geq s \geq 1$

$$\varphi''(\hat{s}) \leq \frac{1}{M},$$

así que

$$\varphi'(s_0) - \varphi'(s) \leq \frac{s_0 - s}{M},$$

es decir

$$\varphi'(s) \geq \varphi'(s_0) + \frac{s - s_0}{M} = q'(s), \quad \forall s \in [1, s_0].$$

Integrando a ambos lados de esta última desigualdad desde s^* hasta s_0

$$\varphi(s_0) - \varphi(s^*) = \int_{s^*}^{s_0} \varphi'(s) ds \geq \int_{s^*}^{s_0} q'(s) ds = q(s_0) - q(s^*) = -q(s^*)$$

pero $\varphi(s^*) \geq 0$ y

$$\begin{aligned} q(s^*) &= (s^* - s_0)\varphi'(s_0) + \frac{(s^* - s_0)^2}{2M} \\ &= -M\varphi'(s_0)\varphi'(s_0) + \frac{[-M\varphi'(s_0)]^2}{2M} \\ &= -\frac{M}{2}[\varphi'(s_0)]^2 \end{aligned}$$

Por lo tanto

$$\varphi(s_0) \geq \varphi(s_0) - \varphi(s^*) \geq \frac{M}{2} [\varphi'(s_0)]^2.$$

□

Proposición 3.6 (inspirada en [3] Proposición 2 (D4)). Sea $\varphi \in \Phi$. Dado $\varepsilon > 0$, existe $\delta > 0$ tal que si $(y_{i_0})^2 \varphi' \left(\frac{x_{i_0}}{y_{i_0}} \right) > \varepsilon$ para algún $i_0 \in \{1, \dots, m\}$, entonces

$$D_\varphi(x, y) > \frac{\delta}{y_{i_0}}.$$

Demostración

Sea $\varepsilon > 0$ cualquiera, y supongamos que

$$(y_{i_0})^2 \varphi' \left(\frac{x_{i_0}}{y_{i_0}} \right) > \varepsilon$$

para algún $i_0 \in \{1, \dots, m\}$.

$$\begin{aligned} (y_{i_0})^2 \varphi' \left(\frac{x_{i_0}}{y_{i_0}} \right) > \varepsilon &\Rightarrow \varphi' \left(\frac{x_{i_0}}{y_{i_0}} \right) > \frac{\varepsilon}{(y_{i_0})^2} > 0 \\ &\Rightarrow \frac{x_{i_0}}{y_{i_0}} > 1, \end{aligned}$$

esta última implicación es válida gracias al observación 3.3.

Por lo tanto, por la proposición anterior

$$\varphi \left(\frac{x_{i_0}}{y_{i_0}} \right) \geq \frac{M}{2} \left[\varphi' \left(\frac{x_{i_0}}{y_{i_0}} \right) \right]^2 > \frac{M\varepsilon^2}{2(y_{i_0})^4}$$

o bien

$$(y_{i_0})^3 \varphi \left(\frac{x_{i_0}}{y_{i_0}} \right) > \frac{M\varepsilon^2}{2y_{i_0}}$$

tomando $\delta = \frac{M\varepsilon^2}{2}$, tenemos que

$$D_\varphi(x, y) = \sum_{i=1}^m (y_i)^3 \varphi \left(\frac{x_i}{y_i} \right) \geq (y_{i_0})^3 \varphi \left(\frac{x_{i_0}}{y_{i_0}} \right) > \frac{\delta}{y_{i_0}}.$$

□

La siguiente proposición es el resultado más importante de todo nuestro trabajo.

Proposición 3.7 (inspirada en [3] **Torema 1**). *Las sucesiones $\{x^k\}$ y $\{\mu^k\}$ dadas en el algoritmo satisfacen que*

- La sucesión $\{[f_i(x^k)]_+\}$ converge a cero para cada $i = 1, \dots, m$.
- La sucesión $\{\mu_i^k f_i(x^k)\}$ converge a cero para todo $i = 1, \dots, m$.
- La sucesión $\{f_0(x^k)\}$ converge a \widehat{f} .
- La sucesión $\{x^k\}$ es acotada.
- Cada punto límite de la sucesión $\{x^k\}$ es solución óptima de (P) y cada punto límite de la sucesión $\{\mu^k\}$ es solución óptima de (D).

Demostración

Para probar (a) procederemos por reducción al absurdo. Supongamos que existe $i_0 \in \{1, \dots, m\}$ tal que la sucesión $\{[f_{i_0}(x^k)]_+\}$ no converge a cero. Así que existen $\varepsilon > 0$ y un conjunto infinito de índices $\{k_j\}$ tales que $f_{i_0}(x^{k_j+1}) > r\varepsilon$.

Pero para $i = 1, \dots, m$

$$f_i(x^{k_j+1}) = r(\mu_i^{k_j})^2 \varphi' \left(\frac{\mu_i^{k_j+1}}{\mu_i^{k_j}} \right),$$

luego

$$\begin{aligned} (f_1(x^{k_j+1}), \dots, f_m(x^{k_j+1}))^T &= \left(r(\mu_1^{k_j})^2 \varphi' \left(\frac{\mu_1^{k_j+1}}{\mu_1^{k_j}} \right), \dots, r(\mu_m^{k_j})^2 \varphi' \left(\frac{\mu_m^{k_j+1}}{\mu_m^{k_j}} \right) \right)^T \\ &= r \nabla_1 D_\varphi(\mu^{k_j+1}, \mu^{k_j}). \end{aligned}$$

Por lo tanto

$$\frac{1}{r} (f_1(x^{k_j+1}), \dots, f_m(x^{k_j+1}))^T = \nabla_1 D_\varphi(\mu^{k_j+1}, \mu^{k_j})$$

y

$$\frac{1}{r} f_{i_0}(x^{k_j+1}) > \varepsilon$$

Por la proposición anterior, existe $\delta > 0$ tal que

$$D_\varphi(\mu^{k_j+1}, \mu^{k_j}) > \frac{\delta}{\mu_{i_0}^{k_j}}.$$

Por otro lado, de (3.2.2)

$$d(\mu^{k_j+1}) - rD_\varphi(\mu^{k_j+1}, \mu^{k_j}) \geq d(\mu^{k_j}) - rD_\varphi(\mu^{k_j}, \mu^{k_j}) = d(\mu^{k_j})$$

o bien

$$d(\mu^{k_j+1}) - d(\mu^{k_j}) \geq rD_\varphi(\mu^{k_j+1}, \mu^{k_j}) > \frac{r\delta}{\mu_{i_0}^{k_j}}.$$

Pero por (3.2.3) se tiene que

$$\frac{r\delta}{\mu_{i_0}^{k_j}} \geq \frac{r\delta}{\widehat{\mu}_{i_0}} > 0,$$

De la proposición 3.3

$$\lim_{j \rightarrow +\infty} [d(\mu^{k_j+1}) - d(\mu^{k_j})] = 0$$

por lo tanto $\frac{r\delta}{\widehat{\mu}_{i_0}} \leq 0$, lo cual es una contradicción. En consecuencia

$$[f_i(x^k)]_+ \rightarrow 0, \quad \forall i = 1, \dots, m.$$

Pasemos ahora a probar (b). Procedamos nuevamente por reducción al absurdo, supongamos que existen $i_0 \in \{1, \dots, m\}$, $\varepsilon > 0$ y un conjunto infinito de índices $\{k_j\}$ tales que

$$|\mu_{i_0}^{k_j+1} f_{i_0}(x^{k_j+1})| \geq \varepsilon, \quad \forall j$$

Por (3.2.3)

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

Pero por (a) $f_{i_0}(x^{k_j+1}) \geq \frac{\varepsilon}{\widehat{\mu}_{i_0}}$ puede ser válida sólo para una cantidad finita de índices k_j , así que podemos suponer, sin perder generalidad, que

$$f_{i_0}(x^{k_j+1}) \leq -\frac{\varepsilon}{\widehat{\mu}_{i_0}}. \quad (3.2.4)$$

para todo j .

Por (3.1.6) y debido a la concavidad de d , se tiene que

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

por (3.1.2) se tiene que para cada $i \in \{1, \dots, m\}$

$$\begin{aligned} f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) &= \mu_i^{k_j} f_i(x^{k_j+1}) \left(\frac{\mu_i^{k_j+1}}{\mu_i^{k_j}} - 1 \right) \\ &= \mu_i^{k_j} f_i(x^{k_j+1}) \left[\theta' \left(\frac{f_i(x^{k_j+1})}{r(\mu_i^{k_j})^2} \right) - 1 \right] \\ &= \mu_i^{k_j} f_i(x^{k_j+1}) \theta' \left(\frac{f_i(x^{k_j+1})}{r(\mu_i^{k_j})^2} \right) - \mu_i^{k_j} f_i(x^{k_j+1}) \\ &= r(\mu_i^{k_j})^3 \left[\frac{f_i(x^{k_j+1})}{r(\mu_i^{k_j})^2} \theta' \left(\frac{f_i(x^{k_j+1})}{r(\mu_i^{k_j})^2} \right) \right] - \mu_i^{k_j} f_i(x^{k_j+1}), \end{aligned}$$

y por la observación 3.1

$$\begin{aligned} f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) &\geq r(\mu_i^{k_j})^3 \left[\frac{f_i(x^{k_j+1})}{r(\mu_i^{k_j})^2} \right] - \mu_i^{k_j} f_i(x^{k_j+1}) \\ &= \mu_i^{k_j} f_i(x^{k_j+1}) - \mu_i^{k_j} f_i(x^{k_j+1}) = 0. \end{aligned}$$

Así que

$$0 \leq \sum_{i=1}^m f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \leq d(\mu^{k_j+1}) - d(\mu^{k_j}),$$

además, por la proposición 3.3 $\lim_{j \rightarrow +\infty} [d(\mu^{k_j+1}) - d(\mu^{k_j})] = 0$, por lo tanto

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

y como $f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) \geq 0$ para todo $i = 1, \dots, m$, entonces

$$\lim_{j \rightarrow +\infty} f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = 0, \quad \forall i = 1, \dots, m. \quad (3.2.5)$$

Pero

$$f_i(x^{k_j+1})(\mu_i^{k_j+1} - \mu_i^{k_j}) = \mu_i^{k_j} f_i(x^{k_j+1}) \left[\theta' \left(\frac{f_i(x^{k_j+1})}{r(\mu_i^{k_j})^2} \right) - 1 \right], \quad \forall i = 1, \dots, m.$$

Así que

$$\lim_{j \rightarrow +\infty} \mu_i^{k_j} f_i(x^{k_j+1}) \left[\theta' \left(\frac{f_i(x^{k_j+1})}{r(\mu_i^{k_j})^2} \right) - 1 \right] = 0, \quad \forall i = 1, \dots, m. \quad (3.2.6)$$

Por otro lado, usando (3.2.3) y (3.2.4), se tiene que

$$\frac{f_{i_0}(x^{k_j+1})}{r(\mu_{i_0}^{k_j})^2} \leq \frac{-\varepsilon}{r(\widehat{\mu}_{i_0})^3} < 0.$$

Por $(\theta 1)$, $(\theta 2)$ y $(\theta 3)$, se tiene que

$$0 < \theta' \left(\frac{f_{i_0}(x^{k_j+1})}{r(\mu_{i_0}^{k_j})^2} \right) \leq \theta' \left(\frac{-\varepsilon}{r(\widehat{\mu}_{i_0})^3} \right) < \theta'(0) = 1.$$

Y por (3.2.6)

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

Esto último junto con (3.2.5) garantiza que

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

lo cual contradice nuestra suposición inicial. En consecuencia

$$\mu_i^k f_i(x^k) \rightarrow 0, \quad \forall i = 1, \dots, m.$$

Para probar (c) notemos que (a) nos dice que la sucesión $\{x^k\}$ es asintóticamente factible, así que dado $\varepsilon > 0$

$$f_0(x^k) \geq \widehat{f} - \varepsilon$$

para k suficientemente grande.

Por la observación 2.2, se tiene que

$$\widehat{f} = \widehat{d} \geq d(\mu^k) = \inf_x \{l(x, \mu^k)\} = l(x^k, \mu^k) = f_0(x^k) + \sum_{i=1}^m \mu_i^k f_i(x^k). \quad (3.2.7)$$

Usando (b) se tiene que para todo $\varepsilon > 0$

$$\widehat{f} - \varepsilon \leq f_0(x^k) \leq \widehat{f} - \sum_{i=1}^m \mu_i^k f_i(x^k) < \widehat{f} + \varepsilon$$

para k suficientemente grande.

Por lo tanto

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

Seguidamente probaremos (d) y (e). De (a) y (b), dado $\varepsilon > 0$

$$f_i(x^k) < \varepsilon, f_0(x^k) < \widehat{f} + \varepsilon, \forall i \in \{1, \dots, m\}. \quad (3.2.8)$$

para k suficientemente grande.

Por (A1), para cualesquiera $\alpha_0, \alpha_1, \dots, \alpha_m$, el conjunto

$$\{x \in \mathbb{R}^n : f_i(x) \leq \alpha_i, \forall i \in \{0, 1, \dots, m\}\}$$

es compacto (ver [5], Corolario 20) y por lo tanto acotado.

Tomando $\alpha_0 = \widehat{f} + 2\varepsilon$ y $\alpha_1 = \dots = \alpha_m = 2\varepsilon$ se tiene que el conjunto

$$\Gamma = \{x \in \mathbb{R}^n : f_i(x) < \varepsilon, f_0(x) < \widehat{f} + \varepsilon, \forall i \in \{1, \dots, m\}\}$$

es acotado.

Pero de (3.2.8) $x^k \in \Gamma$ para k suficientemente grande, es decir, la sucesión $\{x^k\}$ es acotada.

Finalmente, sea $(\bar{x}, \bar{\mu})$ un punto límite de la sucesión $\{(x^k, \mu^k)\}$. Por (a) y (c) tenemos que \bar{x} es una solución óptima de (P). Usando esto último junto con (3.2.7) y (b), se obtiene que $d(\bar{\mu}) = \widehat{f} = \widehat{d}$. Luego $\bar{\mu}$ es una solución óptima de (D). \square

3.3. Implementación del Método

Con el ánimo de estudiar el desempeño del método propuesto anteriormente, presentaremos a continuación un programa en MATLAB que usaremos para su implementación numérica. Usaremos como función de penalidad la función

$$\theta(t) = \begin{cases} -\frac{1}{4t} & \text{si } t < -1/2 \\ 1+t & \text{si } t \geq -1/2 \end{cases}$$

sin embargo, podemos hacer una pequeña variación en el programa principal para así variar la función de penalidad. Además, en cada iteración haremos una actualización de r , haciendo $r^{k+1} = \frac{r^k}{2}$. Las aproximaciones que daremos serán de 4 cifras decimales exactas.

Programa Principal:

```
% Metodo de multiplicadores con multiplicadores cubicos format
short clear all global y2 y1 objetivo epsibfgs epsigold epsilon n
m restrito teta ...
restric delta ...

kmax ifeval igradobjetivo ibfgslanau igolden ipenalty iconstr p
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);

k=0;
xk=x0;
rk=r0;

muk=mu0;
if (n==2)& (m==2),
disp('k x(1) x(2) muk(1) muk(2) fobjetivo flagrangeana'),
fprintf(' %4.0f %14.8f %14.8f %14.8f %14.8f %14.8f %14.8f \n',
k,xk(1),xk(2),muk(1),muk(2),feval('f',xk),feval('flanau',xk,rk,muk));
plot(x0(1),x0(2),'r.')
hold on
end;
```

```

goon=1;
while goon==1,
k=k+1;
y1=f(xk);
xk1=bfqslanau(xk,rk,muk);
y2=f(xk1);
yk1=feval(restric,xk1);
irestric=irestric+1;
for i=1:m, %Actalizacion del multiplicador
if norm(muk(i))<.000000001,
muk1(i)=.000000001;
else
muk1(i)=muk(i)*tetakg(yk1(i)/((muk(i))^(p-1)*rk(i)));
end
end;
% regla de parada: test KKT .
if (norm(muk1-muk)<.00001 )| (k==kmax)| (norm(y2-y1)<.00001), goon=0; end;
rk=rk/2;
xk=xk1;
muk=muk1;
if (n==2)& (m==2),
plot(xk1(1),xk1(2),'r.')
hold on
fprintf(' %4.0f %14.8f %14.8f %14.8f %14.8f %14.8f %14.8f \n',
k,xk1(1),xk1(2),muk1(1),muk1(2),feval('f',xk),feval('flanau',xk,rk,muk));
hold on

```

```

end;
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 búsquedas lineales %8.0f \n', igolden);
fprintf('Nro de iteraciones del BFGS %8.0f \n', ibfgslanau);
zoom on

```

Seguidamente presentaremos algunos ejemplos en los cuales resolveremos un problema de minimización convexo usando el programa anterior, las hipótesis generales de nuestro algoritmo son satisfechas en cada uno de los problemas en cuestión. Cabe destacar que la mayor parte de estos problemas fueron tomados de [2] y algunos fueron modificados para nuestro propósito.

Ejemplo 3.1.

$$\begin{aligned}
 & \text{minimizar} && f(x_1, x_2) = 0,77(x_1 - 2)^2 + (x_2 - 1)^2 \\
 & \text{sujeto a} && x_1^2 - x_2 \leq 0 \\
 & && x_1 + x_2 - 2 \leq 0.
 \end{aligned}$$

Solución

En el Cuadro II mostramos las sucesiones que se generan al aplicar el método, obteniendo que el minimizador aproximado para el problema dado es

$$(x_1^*; x_2^*) = (1,00000057; 0,99999829),$$

el multiplicador óptimo aproximado es

$$(\mu_1^*; \mu_2^*) = (0,51559029; 0,51080927).$$

y estos son encontrados en 10 iteraciones, el mínimo aproximado de la función es

$$f(x_1^*, x_2^*) = 0,76999912.$$

En la figura 3.1 se representan las restricciones del problema y las sucesión x^k , cerca de (x_1^*, x_2^*) .

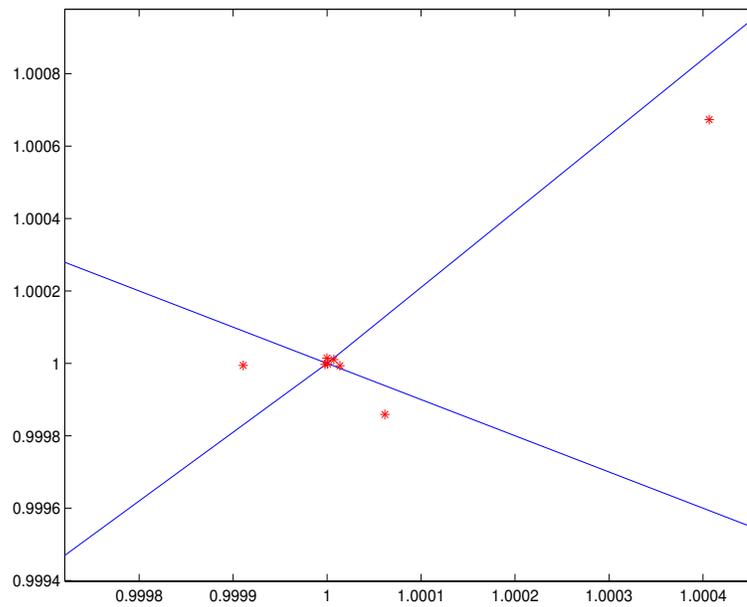


Figura 3.1: Representación gráfica de la sucesión x^k junto con las restricciones. Ejemplo 3.1.

Número de evaluaciones de la función	484
Número de evaluaciones del Gradiente	30
Número de búsquedas lineales	20
Número de iteraciones del BFGS	10

Cuadro I: Número de evaluaciones de las funciones y de algunas de las subrutinas. Ejemplo 3.1.

Ejemplo 3.2.

$$\text{minimizar } f(x_1, x_2) = (x_1 - 5)^2 + (x_2 - 3)^2$$

$$\text{sujeto a } x_1 + x_2 - 3 \leq 0$$

$$-x_1 + 2x_2 - 4 \leq 0.$$

k	x_1^k	x_2^k	μ_1^k	μ_2^k	$f(x^k)$	$L_{r^k}(x^k, \mu^k)$
0	0,00000000	0,00000000	0,16666667	0,16666667	4,08000000	4,07251179
1	1,03432088	1,01523801	0,49415676	0,46402002	0,71828504	0,77957509
2	1,00453467	1,00459039	0,51236764	0,50335047	0,76305351	0,77036190
3	1,00040655	1,00067337	0,51345974	0,51193226	0,76937450	0,77000843
4	1,00006152	0,99985883	0,51757632	0,51068766	0,76990528	0,77000253
5	0,99991074	0,99999437	0,51223182	0,50771477	0,77013747	0,77000195
6	1,00001364	0,99999353	0,51434011	0,50816663	0,76997900	0,77000007
7	0,99999999	1,00001446	0,51253740	0,50998685	0,77000002	0,77000002
8	1,00000716	1,00001050	0,51349024	0,51441927	0,76998898	0,77000010
9	0,99999784	0,99999719	0,51273636	0,51194459	0,77000333	0,77000002
10	1,00000057	0,99999829	0,51559029	0,51080927	0,76999912	0,77000002

Cuadro II: Sucesiones x^k , μ^k , $f(x^k)$ y $L_{r^k}(x^k, \mu^k)$. Ejemplo 3.1.**Solución**

Nótese que aparentemente los valores de μ_2^k en el Cuadro III, son iguales a cero a partir $k = 3$, en realidad son valores muy cercanos a cero, más todos son distintos de cero, los valores en el Cuadro han sido aproximados a 8 cifras decimales.

Minimizador aproximado

$$(x_1^*; x_2^*) = (2,50004157; 0,49995831).$$

Multiplicador óptimo aproximado

$$(\mu_1^*; \mu_2^*) = (4,99993065; 0,00000000).$$

Mínimo aproximado

$$f(x_1^*; x_2^*) = 12,50000058.$$

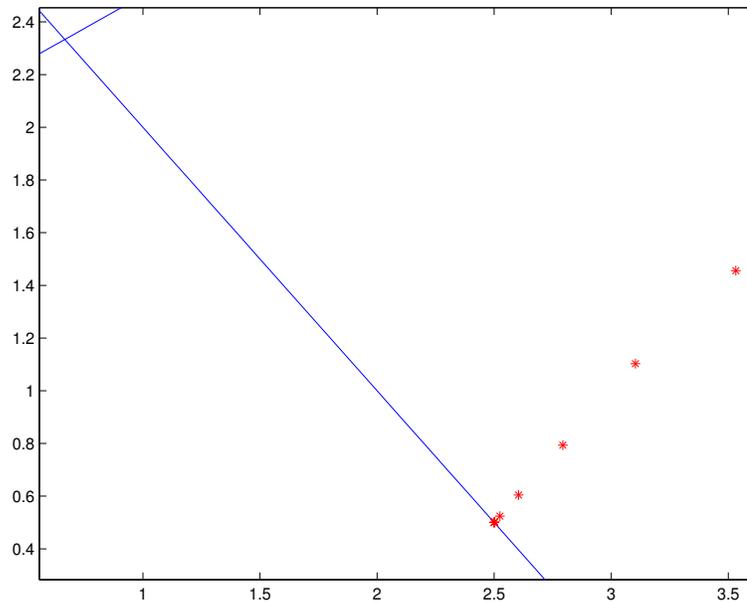
En la figura 3.2 sólo se ve una pequeña parte de la gráfica de la segunda restricción, esto se debe a que el minimizador está sobre la gráfica de la primera restricción (fíjese en la columna correspondiente a μ_2^k de el Cuadro III).

k	x_1^k	x_2^k	μ_1^k	μ_2^k	$f(x^k)$	$L_{r^k}(x^k, \mu^k)$
0	0,00000000	0,00000000	1,00000000	1,00000000	34,00000000	32,28219975
1	3,53190126	1,45557458	2,98747584	0,05410375	4,54056380	11,80010348
2	3,10327555	1,10283912	3,79492315	0,00000404	7,19678305	12,54055800
3	2,79323200	0,79345606	4,41331572	0,00000000	9,73866113	12,63986841
4	2,60436276	0,60421829	4,79140976	0,00000000	11,47884779	12,55088516
5	2,52448743	0,52437609	4,95458019	0,00000000	12,25687620	12,50668496
6	2,50269853	0,50260239	4,98881709	0,00000000	12,47350945	12,50013502
7	2,50057465	0,50048508	5,00241210	0,00000000	12,49470189	12,50001749
8	2,49999310	0,49990692	4,99985373	0,00000000	12,50049993	12,50000028
9	2,50004314	0,49995859	4,99994241	0,00000000	12,49999134	12,50000000
10	2,50004157	0,49995831	4,99993065	0,00000000	12,50000058	12,50000000

Cuadro III: Sucesiones x^k , μ^k , $f(x^k)$ y $L_{r^k}(x^k, \mu^k)$. Ejemplo 3.2.

Número de evaluaciones de la función	402
Número de evaluaciones del Gradiente	25
Número de búsquedas lineales	15
Número de iteraciones del BFGS	10

Cuadro IV: Número de evaluaciones de las funciones y de algunas de las subrutinas. Ejemplo 3.2.

Figura 3.2: Representación gráfica de la sucesión x^k junto con las restricciones. Ejemplo 3.2.

Ejemplo 3.3.

$$\begin{aligned} \text{minimizar } f(x_1, x_2) &= 2x_1^2 + x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \\ \text{sujeto a } x_1 + x_2 - 8 &\leq 0 \\ -x_1 + 2x_2 - 10 &\leq 0. \end{aligned}$$

Solución

Minimizador aproximado

$$(x_1^*; x_2^*) = (2,99990967; 5,00008896).$$

Multiplicador óptimo aproximado

$$(\mu_1^*; \mu_2^*) = (1,99983974; 0,00000000).$$

Mínimo aproximado

$$f(x_1^*, x_2^*) = -28,99999723.$$

k	x_1^k	x_2^k	μ_1^k	μ_2^k	$f(x^k)$	$L_{r,k}(x^k, \mu^k)$
0	3,00000000	3,00000000	0,33333333	0,33333333	-21,00000000	-21,10573890
1	3,19759604	5,29206809	1,80232570	0,00354291	-29,93135908	-28,91579074
2	3,05236878	5,07855582	1,94760976	0,00000000	-29,25842094	-28,98582858
3	3,00875949	5,01357454	1,99347938	0,00000000	-29,04456814	-28,99904484
4	3,00031717	5,00063141	1,99728609	0,00000000	-29,00189695	-28,99999877
5	2,99998326	5,00025499	1,99919468	0,00000000	-29,00047642	-28,99999966
6	2,99990027	5,00013375	1,99973916	0,00000000	-29,00006797	-28,99999993
7	2,99989476	5,00008955	1,99923695	0,00000000	-28,99996857	-28,99999993
8	2,99991275	5,00010002	2,00005499	0,00000000	-29,00002551	-28,99999995
9	2,99990737	5,00009029	1,99975661	0,00000000	-28,99999530	-28,99999996
10	2,99991139	5,00009166	2,00053931	0,00000000	-29,00000607	-28,99999996
11	2,99990967	5,00008896	1,99983974	0,00000000	-28,99999723	-28,99999996

Cuadro V: Sucesiones x^k , μ^k , $f(x^k)$ y $L_{r,k}(x^k, \mu^k)$. Ejemplo 3.3.

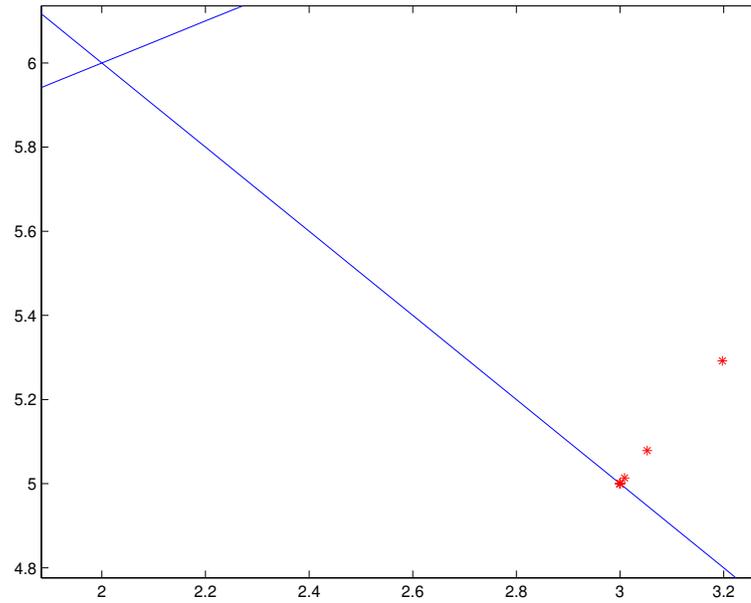


Figura 3.3: Representación gráfica de la sucesión x^k junto con las restricciones. Ejemplo 3.3.

Número de evaluaciones de la función	457
Número de evaluaciones del Gradiente	29
Número de búsquedas lineales	18
Número de iteraciones del BFGS	11

Cuadro VI: Número de evaluaciones de las funciones y de algunas de las subrutinas. Ejemplo 3.3.

Ejemplo 3.4.

$$\text{minimizar } f(x_1, x_2) = e^{x_1} - x_1x_2 + x_2^2$$

$$\text{sujeto a } x_1^2 + x_2^2 - 4 \leq 0$$

$$2x_1 + x_2 - 2 \leq 0.$$

Solución

Minimizador aproximado

$$(x_1^*; x_2^*) = (-1,85550259; -0,74639827).$$

Multiplicador óptimo aproximado

$$(\mu_1^*; \mu_2^*) = (0,24336359; 0,00000000).$$

Mínimo aproximado

$$f(x_1^*, x_2^*) = -0,67145922.$$

k	x_1^k	x_2^k	μ_1^k	μ_2^k	$f(x^k)$	$L_{r,k}(x^k, \mu^k)$
0	0,00000000	0,00000000	0,33333333	0,33333333	1,00000000	0,89944273
1	-1,84871452	-0,74337610	0,24439350	0,00143759	-0,66424274	-0,66823113
2	-1,85546625	-0,74638711	0,24315558	0,00000000	-0,67142236	-0,67145896
3	-1,85551181	-0,74640327	0,24384499	0,00000000	-0,67146935	-0,67145910
4	-1,85549884	-0,74639725	0,24334673	0,00000000	-0,67145546	-0,67145915
5	-1,85550259	-0,74639827	0,24336359	0,00000000	-0,67145922	-0,67145916

Cuadro VII: Sucesiones x^k , μ^k , $f(x^k)$ y $L_{r,k}(x^k, \mu^k)$. Ejemplo 3.4.

Número de evaluaciones de la función	333
Número de evaluaciones del Gradiente	17
Número de búsquedas lineales	12
Número de iteraciones del BFGS	5

Cuadro VIII: Número de evaluaciones de las funciones y de algunas de las subrutinas. Ejemplo 3.4.

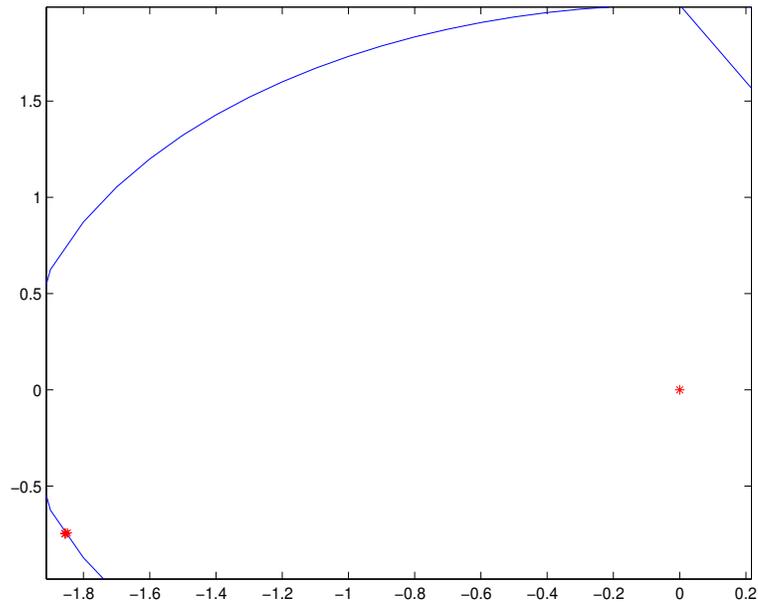


Figura 3.4: Representación gráfica de la sucesión x^k junto con las restricciones. Ejemplo 3.4.

Ejemplo 3.5.

$$\begin{aligned} \text{minimizar} \quad & f(x_1, x_2) = 4x_1^2 - 5x_1x_2 + x_2^2 \\ \text{sujeto a} \quad & x_1^2 - x_2 + 2 \leq 0 \\ & x_1 + x_2 - 6 \leq 0. \end{aligned}$$

Solución

Minimizador aproximado

$$(x_1^*; x_2^*) = (1,56155411; 4,43844741).$$

Multiplicador óptimo aproximado

$$(\mu_1^*; \mu_2^*) = (2,61184120; 1,54270818).$$

Mínimo aproximado

$$f(x_1^*, x_2^*) = -5,20075859.$$

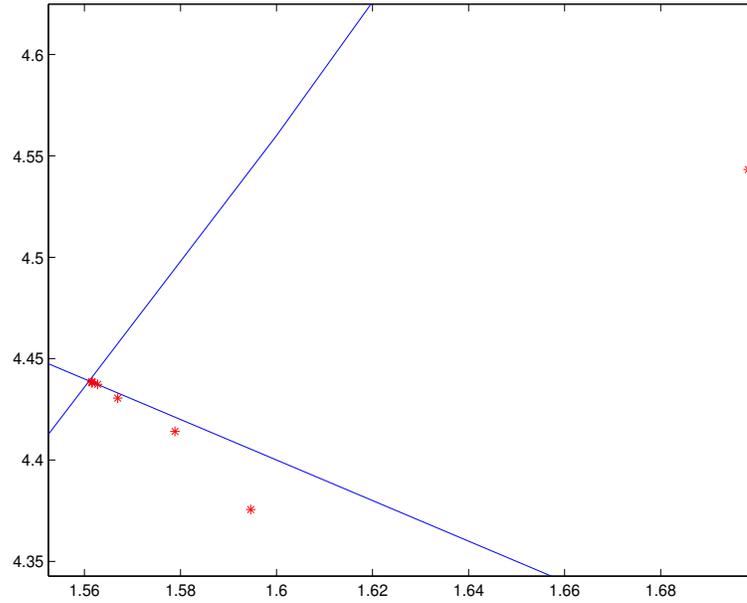


Figura 3.5: Representación gráfica de la sucesión x^k junto con las restricciones. Ejemplo 3.5.

Número de evaluaciones de la función	555
Número de evaluaciones del Gradiente	32
Número de búsquedas lineales	23
Número de iteraciones del BFGS	9

Cuadro IX: Número de evaluaciones de las funciones y de algunas de las subrutinas. Ejemplo 3.5.

k	x_1^k	x_2^k	μ_1^k	μ_2^k	$f(x^k)$	$L_{r,k}(x^k, \mu^k)$
0	0,00000000	0,00000000	0,16666667	0,16666667	0,00000000	12,32457358
1	1,69823227	4,54330187	2,21081258	1,61587152	-6,40034602	-5,16824917
2	1,59463951	4,37559856	2,36213847	1,57903452	-5,57014826	-5,19719935
3	1,57885853	4,41408097	2,49543004	1,56114890	-5,39075910	-5,19529897
4	1,56694110	4,43055798	2,57476364	1,54833310	-5,26105532	-5,19927631
5	1,56275359	4,43733992	2,60495736	1,54929941	-5,21356373	-5,20061665
6	1,56159656	4,43809094	2,61101193	1,54284498	-5,20155126	-5,20074150
7	1,56155433	4,43843485	2,61143046	1,54239588	-5,20077416	-5,20074626
8	1,56155317	4,43844755	2,61146686	1,54245558	-5,20074932	-5,20074627
9	1,56155411	4,43844741	2,61184120	1,54270818	-5,20075859	-5,20074626

Cuadro X: Sucesiones x^k , μ^k , $f(x^k)$ y $L_{r,k}(x^k, \mu^k)$. Ejemplo 3.5.

Ejemplo 3.6.

$$\begin{aligned} \text{minimizar } f(x_1, x_2) &= x_1^2 - 2x_1x_2 + x_2^2 + e^{-x_1-x_2} \\ \text{sujeto a } x_1^2 + x_2^2 - 4 &\leq 0 \\ x_1 + x_2 - 1 &\leq 0. \end{aligned}$$

Solución

Minimizador aproximado

$$(x_1^*; x_2^*) = (0,50000018; 0,50000018).$$

Multiplicador óptimo aproximado

$$(\mu_1^*; \mu_2^*) = (0,00000000; 0,36788246).$$

Mínimo aproximado

$$f(x_1^*, x_2^*) = 0,36787931.$$

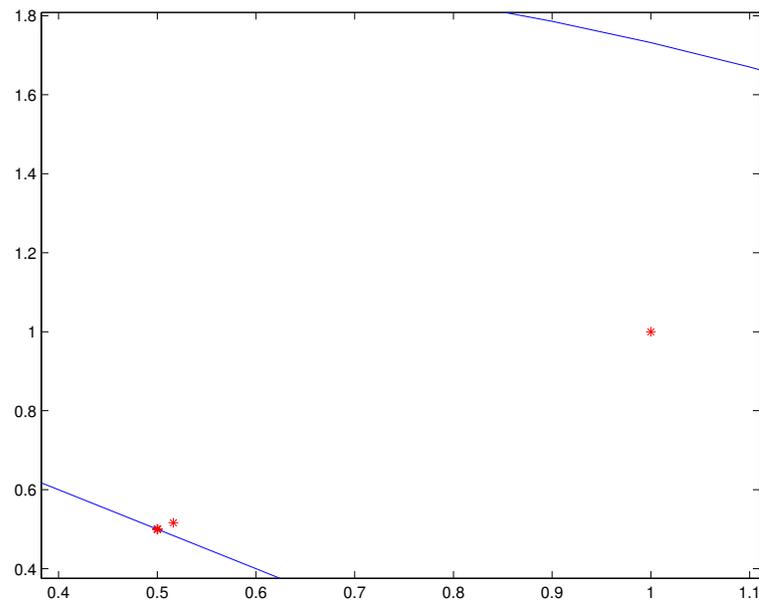


Figura 3.6: Representación gráfica de la sucesión x^k junto con las restricciones. Ejemplo 3.6.

k	x_1^k	x_2^k	μ_1^k	μ_2^k	$f(x^k)$	$L_{r^k}(x^k, \mu^k)$
0	1,00000000	1,00000000	0,16666667	0,16666667	0,13533528	3,30200195
1	0,51629554	0,51629554	0,00008383	0,36221320	0,35608312	0,37082051
2	0,50036664	0,50036664	0,00000000	0,36626206	0,36760978	0,36788129
3	0,50008157	0,50008157	0,00000000	0,36804372	0,36781943	0,36787976
4	0,49995286	0,49995286	0,00000000	0,36599459	0,36791412	0,36787981
5	0,50002305	0,50002305	0,00000000	0,36801001	0,36786248	0,36787954
6	0,49999891	0,49999891	0,00000000	0,36781963	0,36788025	0,36787944
7	0,50000018	0,50000018	0,00000000	0,36788246	0,36787931	0,36787944

Cuadro XI: Sucesiones x^k , μ^k , $f(x^k)$ y $L_{r^k}(x^k, \mu^k)$. Ejemplo 3.6.

Número de evaluaciones de la función	193
Número de evaluaciones del Gradiente	15
Número de búsquedas lineales	8
Número de iteraciones del BFGS	7

Cuadro XII: Número de evaluaciones de las funciones y de algunas de las subrutinas. Ejemplo 3.6.

Obsérvese que el valor óptimo aproximado de la función lagrangeano, en todos los ejemplos presentados, es bastante próximo al mínimo aproximado de la función objetivo.

Algo más que puede observarse, es que el número de evaluaciones de la función objetivo y de los gradientes, es relativamente bajo, además, la convergencia del método, con una tolerancia de cuatro cifras decimales exactas, se obtiene con una cantidad de iteraciones bastante pequeña.

Es de hacer notar que los ejemplos presentados aca, fueron escogidos de tal manera que fuesen todos de dos variables y dos restricciones, principalmente para poder graficar las restricciones, sin embargo, se hicieron pruebas del método con mayor cantidad de variables y mayor número de restricciones, en los cuales se obtuvieron resultados tan buenos como los obtenidos en estos ejemplos.

Conclusiones

Se ha realizado un estudio de los métodos de multiplicadores que conducen, vía dualidad de Fenchel, a métodos de punto proximal con cuasidistancias del tipo φ -divergencias. Al respecto, se ha introducido un nuevo método de multiplicadores que contiene potencias cúbicas en los multiplicadores que definen la función lagrangeano aumentado. Esto trae como consecuencia el realizar un estudio tanto algorítmico como analítico del método. Se han obtenido resultados de convergencia similares a los obtenidos para el caso con multiplicadores con potencias lineales y/o cuadráticas de los casos ya conocidos, mencionando por ejemplo, la acotabilidad de la sucesión de iterados primal, todo punto límite de esta es solución óptima de (P) . En el estudio computacional del método, se obtuvieron resultados alentadores en el sentido de que se obtuvieron las respuestas correctas a los problemas considerados, incluso en algunos con multiplicadores iguales a cero en la solución óptima.

Como futuro trabajo que debe continuar al presente, se plantea la interrogante de estudiar métodos con potencias reales en los multiplicadores en la función objetivo, así como también el de efectuar un estudio comparativo en estas metodologías para estimar la mejor potencia en algún sentido, por ejemplo, estudiando la rapidez de convergencia.

Bibliografia

- [1] Alfred Auslender, Marc Teboulle and Sami Ben-Tiba, *Interior Proximal and Multiplier Methods Based on Second Order Homogeneous Kernels*, Mathematics of Operations Research, Vol. 24, N° 3, (1999), 645-668.
- [2] Mokhtar S. Bazaraa and C. M. Shetty, *Nonlinear Programming. Theory and Algorithms*, Jhon Wiley and Sons, (1979).
- [3] Aharon Ben-Tal and Michael Zibulevsky, *Penalty/Barrier Multiplier Methods*, Society for Industrial and Applied Mathematics, 7, (1997), 346-366.
- [4] Castillo Rómulo, *Métodos de Lagrangeano Aumentado Usando Funções de Penalidades Generalizadas para Programação não Linear*, COPPE, UFRJ, Brasil, (1998).
- [5] Anthony V. Fiacco and Garth P. McCormick, *Nonlinear Programming. Sequential Unconstrained Minimization Techniques*, SIAM, (1990).
- [6] Jean B. Hiriart-Urruty and Lemarréchal C., *Convex Analysis and Minimization Algorithms II*, New York, Springer-Verlag, (1993).
- [7] Alfredo Iusem, *Métodos de Ponto Proximal em Otimização*, 20° Colóquio Brasileiro de Matemática, IMPA, 24-28 julho, (1995).
- [8] Alfredo N. Iusem, B. F. Svaiter and Marc Teboulle, *Entropy-Like Proximal Methods in Convex Programming*, Mathematics of Operations Research, Vol. 19, N° 4, (1994), 790-814.
- [9] Roman Polyak and Marc Teboulle, *Nonlinear Rescaling and Proximal-Like Methods in Convex Optimization*, Mathematical Programming, 76, (1997), 265-284.

-
- [10] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N. J. (1970).